P1 1175593

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office

May 27, 2004

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *60/528,750*
FILING DATE: *December 10, 2003*
RELATED PCT APPLICATION NUMBER: *PCT/US04/09215*

REC'D 01 JUN 2004
WIPO PCT

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

M. Sias

M. SIAS
Certifying Officer

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

BEST AVAILABLE COPY

Atty. Docket No.: PF-1724 P	Type a plus sign (+) inside this box ➡	+

121003 14202 U.S. PTO

Certificate of Mailing

Certificate of Express Mail" mailing label number EV 310 863 595 US I hereby certify that this document and referenced attachments are being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. § 1.10, addressed to: Mail Stop Provisional Patent Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on December 10, 2003.

By: Nancy Ramos Printed: Nancy Ramos

22581 U.S. PTO 60/528750 121003

PROVISIONAL APPLICATION FOR PATENT COVER SHEET
(Large Entity)

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(b)(2).

TITLE OF THE INVENTION (500 characters max)	
KINASES AND PHOSPHATASES	
INVENTOR(S)/APPLICANT(S)	
FIRST NAME/MIDDLE INITIAL/LAST NAME	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
Narinder K. Chawla	Union City, California
Shanya D. Becha	San Francisco, California
Amy D. Wilson	Encino, California
Pei Jin	Palo Alto, California

CORRESPONDENCE ADDRESS (Customer No.: 27904)	
INCYTE CORPORATION Legal Department 3160 Porter Drive Palo Alto, California 94304	Phone: (650) 855-0555 Fax: (650) 849-8886 or (650) 845-4166

Atty. Docket No.: PF-1724 P	Type a plus sign (+) inside this box ➡	+

ENCLOSURES

Enclosed are:

1. Return Receipt Postcard;
2. Provisional Application Cover Sheet w/certificate of Express Mailing (2 pp., in duplicate);
3. 99 Pages of Specification (1 - 99) ;
4. 11 Pages of Claims (100 - 110) ;
5. 1 Page of Abstract (111) ;
6. 45 Pages of Tables (Tables 1 - 8) ; and
7. 37 Pages of Sequence Listing (1 - 37) .

METHOD OF PAYMENT

Applicants hereby authorize the Commissioner to treat any concurrent or future reply in this application that requires a petition for an extension of time under 37 CFR 1.136(a) to be timely, as incorporating a petition for extension of time for the appropriate length of time; and to charge all required fees, including fees under 37 CFR 1.16, 1.17, and all required extension of time fees, or to credit any overpayment, to Deposit Account **09-0108**. **This sheet is enclosed in duplicate.**

PROVISIONAL FILING FEE AMOUNT: $ 160.00

The invention was not made by an agency of the United States Government or under a contract with an agency of the United States Government.

Respectfully submitted,

INCYTE CORPORATION

Date: December 10, 2003

Barrie Greene

Barrie D. Greene
Reg. No. 46,740
Direct Dial Telephone: (650) 621-7576

EXPRESS MAIL NO. EV310863595US

KINASES AND PHOSPHATASES

TECHNICAL FIELD

The invention relates to novel nucleic acids, kinases and phosphatases encoded by these nucleic acids, and to the use of these nucleic acids and proteins in the diagnosis, treatment, and prevention of cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers. The invention also relates to the assessment of the effects of exogenous compounds on the expression of nucleic acids and kinases and phosphatases.

BACKGROUND OF THE INVENTION

Reversible protein phosphorylation is the ubiquitous strategy used to control many of the intracellular events in eukaryotic cells. It is estimated that more than ten percent of proteins active in a typical mammalian cell are phosphorylated. Kinases catalyze the transfer of high-energy phosphate groups from adenosine triphosphate (ATP) to target proteins on the hydroxyamino acid residues serine, threonine, or tyrosine. Phosphatases, in contrast, remove these phosphate groups. Extracellular signals including hormones, neurotransmitters, and growth and differentiation factors can activate kinases, which can occur as cell surface receptors or as the activator of the final effector protein, as well as other locations along the signal transduction pathway. Cascades of kinases occur, as well as kinases sensitive to second messenger molecules. This system allows for the amplification of weak signals (low abundance growth factor molecules, for example), as well as the synthesis of many weak signals into an all-or-nothing response. Phosphatases, then, are essential in determining the extent of phosphorylation in the cell and, together with kinases, regulate key cellular processes such as metabolic enzyme activity, proliferation, cell growth and differentiation, cell adhesion, and cell cycle progression.

KINASES

Kinases comprise the largest known enzyme superfamily and vary widely in their target molecules. Kinases catalyze the transfer of high energy phosphate groups from a phosphate donor to a phosphate acceptor. Nucleotides usually serve as the phosphate donor in these reactions, with most kinases utilizing adenosine triphosphate (ATP). The phosphate acceptor can be any of a variety of molecules, including nucleosides, nucleotides, lipids, carbohydrates, and proteins. Proteins are phosphorylated on hydroxyamino acids. Addition of a phosphate group alters the local charge on the acceptor molecule, causing internal conformational changes and potentially influencing intermolecular contacts. Reversible protein phosphorylation is the primary method for regulating protein activity in eukaryotic cells. In general, proteins are activated by phosphorylation in response

to extracellular signals such as hormones, neurotransmitters, and growth and differentiation factors. The activated proteins initiate the cell's intracellular response by way of intracellular signaling pathways and second messenger molecules such as cyclic nucleotides, calcium-calmodulin, inositol, and various mitogens, that regulate protein phosphorylation.

Kinases are involved in all aspects of a cell's function, from basic metabolic processes, such as glycolysis, to cell-cycle regulation, differentiation, and communication with the extracellular environment through signal transduction cascades. Inappropriate phosphorylation of proteins in cells has been linked to changes in cell cycle progression and cell differentiation. Changes in the cell cycle have been linked to induction of apoptosis or cancer. Changes in cell differentiation have been linked to diseases and disorders of the reproductive system, immune system, and skeletal muscle.

There are two classes of protein kinases. One class, protein tyrosine kinases (PTKs), phosphorylates tyrosine residues, and the other class, protein serine/threonine kinases (STKs), phosphorylates serine and threonine residues. Some PTKs and STKs possess structural characteristics of both families and have dual specificity for both tyrosine and serine/threonine residues. Almost all kinases contain a conserved 250-300 amino acid catalytic domain containing specific residues and sequence motifs characteristic of the kinase family. The protein kinase catalytic domain can be further divided into 11 subdomains. N-terminal subdomains I-IV fold into a two-lobed structure which binds and orients the ATP donor molecule, and subdomain V spans the two lobes. C-terminal subdomains VI-XI bind the protein substrate and transfer the gamma phosphate from ATP to the hydroxyl group of a tyrosine, serine, or threonine residue. Each of the 11 subdomains contains specific catalytic residues or amino acid motifs characteristic of that subdomain. For example, subdomain I contains an 8-amino acid glycine-rich ATP binding consensus motif, subdomain II contains a critical lysine residue required for maximal catalytic activity, and subdomains VI through IX comprise the highly conserved catalytic core. PTKs and STKs also contain distinct sequence motifs in subdomains VI and VIII which may confer hydroxyamino acid specificity.

In addition, kinases may also be classified by additional amino acid sequences, generally between 5 and 100 residues, which either flank or occur within the kinase domain. These additional amino acid sequences regulate kinase activity and determine substrate specificity. (Reviewed in Hardie, G. and S. Hanks (1995) The Protein Kinase Facts Book, Vol I, pp. 17-20 Academic Press, San Diego CA.). In particular, two protein kinase signature sequences have been identified in the kinase domain, the first containing an active site lysine residue involved in ATP binding, and the second containing an aspartate residue important for catalytic activity. If a protein analyzed includes the two protein kinase signatures, the probability of that protein being a protein kinase is close to 100% (PROSITE: PDOC00100, November 1995).

Protein Tyrosine Kinases

Protein tyrosine kinases (PTKs) may be classified as either transmembrane, receptor PTKs or nontransmembrane, nonreceptor PTK proteins. Transmembrane tyrosine kinases function as receptors for most growth factors. Growth factors bind to the receptor tyrosine kinase (RTK), which causes the receptor to phosphorylate itself (autophosphorylation) and specific intracellular second messenger proteins. Growth factors (GF) that associate with receptor PTKs include epidermal GF, platelet-derived GF, fibroblast GF, hepatocyte GF, insulin and insulin-like GFs, nerve GF, vascular endothelial GF, and macrophage colony stimulating factor.

Nontransmembrane, nonreceptor PTKs lack transmembrane regions and, instead, form signaling complexes with the cytosolic domains of plasma membrane receptors. Receptors that function through non-receptor PTKs include those for cytokines and hormones (growth hormone and prolactin), and antigen-specific receptors on T and B lymphocytes.

Many PTKs were first identified as oncogene products in cancer cells in which PTK activation was no longer subject to normal cellular controls. In fact, about one third of the known oncogenes encode PTKs. Furthermore, cellular transformation (oncogenesis) is often accompanied by increased tyrosine phosphorylation activity (Charbonneau, H. and N.K. Tonks (1992) Annu. Rev. Cell Biol. 8:463-493). Regulation of PTK activity may therefore be an important strategy in controlling some types of cancer.

Protein Serine/Threonine Kinases

Protein serine/threonine kinases (STKs) are nontransmembrane proteins. A subclass of STKs are known as ERKs (extracellular signal regulated kinases) or MAPs (mitogen-activated protein kinases) and are activated after cell stimulation by a variety of hormones and growth factors. Cell stimulation induces a signaling cascade leading to phosphorylation of MEK (MAP/ERK kinase) which, in turn, activates ERK via serine and threonine phosphorylation. A varied number of proteins represent the downstream effectors for the active ERK and implicate it in the control of cell proliferation and differentiation, as well as regulation of the cytoskeleton. Activation of ERK is normally transient, and cells possess dual specificity phosphatases that are responsible for its down-regulation. Also, numerous studies have shown that elevated ERK activity is associated with some cancers. Other STKs include the second messenger dependent protein kinases such as the cyclic-AMP dependent protein kinases (PKA), calcium-calmodulin (CaM) dependent protein kinases, and the mitogen-activated protein kinases (MAP); the cyclin-dependent protein kinases; checkpoint and cell cycle kinases; Numb-associated kinase (Nak); human Fused (hFu); proliferation-related kinases; 5'-AMP-activated protein kinases; and kinases involved in apoptosis.

One member of the ERK family of MAP kinases, ERK 7, is a novel 61-kDa protein that has motif similarities to ERK1 and ERK2, but is not activated by extracellular stimuli as are ERK1 and

ERK2 nor by the common activators, c-Jun N-terminal kinase (JNK) and p38 kinase. ERK7 regulates its nuclear localization and inhibition of growth through its C-terminal tail, not through the kinase domain as is typical with other MAP kinases (Abe, M.K. (1999) Mol. Cell. Biol. 19:1301-1312).

The second messenger dependent protein kinases primarily mediate the effects of second messengers such as cyclic AMP (cAMP), cyclic GMP, inositol triphosphate, phosphatidylinositol, 3,4,5-triphosphate, cyclic ADP ribose, arachidonic acid, diacylglycerol and calcium-calmodulin. The PKAs are involved in mediating hormone-induced cellular responses and are activated by cAMP produced within the cell in response to hormone stimulation. cAMP is an intracellular mediator of hormone action in all animal cells that have been studied. Hormone-induced cellular responses include thyroid hormone secretion, cortisol secretion, progesterone secretion, glycogen breakdown, bone resorption, and regulation of heart rate and force of heart muscle contraction. PKA is found in all animal cells and is thought to account for the effects of cAMP in most of these cells. Altered PKA expression is implicated in a variety of disorders and diseases including cancer, thyroid disorders, diabetes, atherosclerosis, and cardiovascular disease (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, New York NY, pp. 416-431, 1887).

The casein kinase I (CKI) gene family is another subfamily of serine/threonine protein kinases. This continuously expanding group of kinases have been implicated in the regulation of numerous cytoplasmic and nuclear processes, including cell metabolism and DNA replication and repair. CKI enzymes are present in the membranes, nucleus, cytoplasm and cytoskeleton of eukaryotic cells, and on the mitotic spindles of mammalian cells (Fish, K.J. et al. (1995) J. Biol. Chem. 270:14875-14883).

The CKI family members all have a short amino-terminal domain of 9-76 amino acids, a highly conserved kinase domain of 284 amino acids, and a variable carboxyl-terminal domain that ranges from 24 to over 200 amino acids in length (Cegielska, A. et al. (1998) J. Biol. Chem. 273:1357-1364). The CKI family is comprised of highly related proteins, as seen by the identification of isoforms of casein kinase I from a variety of sources. There are at least five mammalian isoforms, α, β, γ, δ, and ε. Fish et al. identified CKI-epsilon from a human placenta cDNA library. It is a basic protein of 416 amino acids and is closest to CKI-delta. Through recombinant expression, it was determined to phosphorylate known CKI substrates and was inhibited by the CKI-specific inhibitor CKI-7. The human gene for CKI-epsilon was able to rescue yeast with a slow-growth phenotype caused by deletion of the yeast CKI locus, HRR250 (Fish et al., *supra*).

The mammalian circadian mutation tau was found to be a semidominant autosomal allele of CKI-epsilon that markedly shortens period length of circadian rhythms in Syrian hamsters. The tau locus is encoded by casein kinase I-epsilon, which is also a homolog of the *Drosophila* circadian gene double-time. Studies of both the wildtype and tau mutant CKI-epsilon enzyme indicated that the

mutant enzyme has a noticeable reduction in the maximum velocity and autophosphorylation state. Further, *in vitro*, CKI-epsilon is able to interact with mammalian PERIOD proteins, while the mutant enzyme is deficient in its ability to phosphorylate PERIOD. Lowrey et al. have proposed that CKI-epsilon plays a major role in delaying the negative feedback signal within the transcription-translation-based autoregulatory loop that composes the core of the circadian mechanism. Therefore the CKI-epsilon enzyme is an ideal target for pharmaceutical compounds influencing circadian rhythms, jet-lag and sleep, in addition to other physiologic and metabolic processes under circadian regulation (Lowrey, P.L. et al. (2000) Science 288:483-491).

Homeodomain-interacting protein kinases (HIPKs) are serine/threonine kinases and novel members of the DYRK kinase subfamily (Hofmann, T.G. et al. (2000) Biochimie 82:1123-1127). HIPKs contain a conserved protein kinase domain separated from a domain that interacts with homeoproteins. HIPKs are nuclear kinases, and HIPK2 is highly expressed in neuronal tissue (Kim, Y.H. et al. (1998) J. Biol. Chem. 273:25875-25879; Wang, Y. et al. (2001) Biochim. Biophys. Acta 1518:168-172). HIPKs act as corepressors for homeodomian transcription factors. This corepressor activity is seen in posttranslational modifications such as ubiquitination and phosphorylation, each of which are important in the regulation of cellular protein function (Kim, Y.H. et al. (1999) Proc. Natl. Acad. Sci. USA 96:12350-12355).

The human h-warts protein, a homolog of *Drosophila* warts tumor suppressor gene, maps to chromosome 6q24-25.1. It has a serine/threonine kinase domain and is localized to centrosomes in interphase cells. It is involved in mitosis and functions as a component of the mitotic apparatus (Nishiyama, Y. et al. (1999) FEBS Lett. 459:159-165).

<u>Calcium-Calmodulin Dependent Protein Kinases</u>

Calcium-calmodulin dependent (CaM) kinases are involved in regulation of smooth muscle contraction, glycogen breakdown (phosphorylase kinase), and neurotransmission (CaM kinase I and CaM kinase II). CaM dependent protein kinases are activated by calmodulin, an intracellular calcium receptor, in response to the concentration of free calcium in the cell. Many CaM kinases are also activated by phosphorylation. Some CaM kinases are also activated by autophosphorylation or by other regulatory kinases. CaM kinase I phosphorylates a variety of substrates including the neurotransmitter-related proteins synapsin I and II, the gene transcription regulator, CREB, and the cystic fibrosis conductance regulator protein, CFTR (Haribabu, B. et al. (1995) EMBO J. 14:3679-3686). CaM kinase II also phosphorylates synapsin at different sites and controls the synthesis of catecholamines in the brain through phosphorylation and activation of tyrosine hydroxylase. CaM kinase II controls the synthesis of catecholamines and seratonin, through phosphorylation/activation of tyrosine hydroxylase and tryptophan hydroxylase, respectively (Fujisawa, H. (1990) BioEssays 12:27-29). The mRNA encoding a calmodulin-binding protein kinase-like protein was found to be

enriched in mammalian forebrain. This protein is associated with vesicles in both axons and dendrites and accumulates largely postnatally. The amino acid sequence of this protein is similar to CaM-dependent STKs, and the protein binds calmodulin in the presence of calcium (Godbout, M. et al. (1994) J. Neurosci. 14:1-13).

<u>Mitogen-Activated Protein Kinases</u>

The mitogen-activated protein kinases (MAP), which mediate signal transduction from the cell surface to the nucleus via phosphorylation cascades, are another STK family that regulates intracellular signaling pathways. Several subgroups have been identified, and each manifests different substrate specificities and responds to distinct extracellular stimuli (Egan, S.E. and R.A. Weinberg (1993) Nature 365:781-783). There are three kinase modules comprising the MAP kinase cascade: MAPK (MAP), MAPK kinase (MAP2K, MAPKK, or MKK), and MKK kinase (MAP3K, MAPKKK, OR MEKK) (Wang,X.S. et al (1998) Biochem. Biophys. Res. Commun. 253:33-37). The extracellular-regulated kinase (ERK) pathway is activated by growth factors and mitogens, for example, epidermal growth factor (EGF), ultraviolet light, hyperosmolar medium, heat shock, or endotoxic lipopolysaccharide (LPS). The closely related though distinct parallel pathways, the c-Jun N-terminal kinase (JNK), or stress-activated kinase (SAPK) pathway, and the p38 kinase pathway are activated by stress stimuli and proinflammatory cytokines such as tumor necrosis factor (TNF) and interleukin-1 (IL-1). Altered MAP kinase expression is implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development.. MAP kinase signaling pathways are present in mammalian cells as well as in yeast.

<u>Cyclin-Dependent Protein Kinases</u>

The cyclin-dependent protein kinases (CDKs) are STKs that control the progression of cells through the cell cycle. The entry and exit of a cell from mitosis are regulated by the synthesis and destruction of a family of activating proteins called cyclins. Cyclins are small regulatory proteins that bind to and activate CDKs, which then phosphorylate and activate selected proteins involved in the mitotic process. CDKs are unique in that they require multiple inputs to become activated. In addition to cyclin binding, CDK activation requires the phosphorylation of a specific threonine residue and the dephosphorylation of a specific tyrosine residue on the CDK.

Another family of STKs associated with the cell cycle are the NIMA (never in mitosis)-related kinases (Neks). Both CDKs and Neks are involved in duplication, maturation, and separation of the microtubule organizing center, the centrosome, in animal cells (Fry, A.M. et al. (1998) EMBO J. 17:470-481).

<u>Checkpoint and Cell Cycle Kinases</u>

In the process of cell division, the order and timing of cell cycle transitions are under control of cell cycle checkpoints, which ensure that critical events such as DNA replication and chromosome

segregation are carried out with precision. If DNA is damaged, e.g. by radiation, a checkpoint pathway is activated that arrests the cell cycle to provide time for repair. If the damage is extensive, apoptosis is induced. In the absence of such checkpoints, the damaged DNA is inherited by aberrant cells which may cause proliferative disorders such as cancer. Protein kinases play an important role in this process. For example, a specific kinase, checkpoint kinase 1 (Chk1), has been identified in yeast and mammals, and is activated by DNA damage in yeast. Activation of Chk1 leads to the arrest of the cell at the G2/M transition (Sanchez, Y. et al. (1997) Science 277:1497-1501). Specifically, Chk1 phosphorylates the cell division cycle phosphatase CDC25, inhibiting its normal function which is to dephosphorylate and activate the cyclin-dependent kinase Cdc2. Cdc2 activation controls the entry of cells into mitosis (Peng, C.-Y. et al. (1997) Science 277:1501-1505). Thus, activation of Chk1 prevents the damaged cell from entering mitosis. A deficiency in a checkpoint kinase, such as Chk1, may also contribute to cancer by failure to arrest cells with damaged DNA at other checkpoints such as G2/M.

<u>Proliferation-Related Kinases</u>

Proliferation-related kinase is a serum/cytokine inducible STK that is involved in regulation of the cell cycle and cell proliferation in human megakarocytic cells (Li, B. et al. (1996) J. Biol. Chem. 271:19402-19408). Proliferation-related kinase is related to the polo (derived from *Drosophila* polo gene) family of STKs implicated in cell division. Proliferation-related kinase is downregulated in lung tumor tissue and may be a proto-oncogene whose deregulated expression in normal tissue leads to oncogenic transformation.

<u>5'-AMP-activated protein kinase</u>

A ligand-activated STK protein kinase is 5'-AMP-activated protein kinase (AMPK) (Gao, G. et al. (1996) J. Biol Chem. 271:8675-8681). Mammalian AMPK is a regulator of fatty acid and sterol synthesis through phosphorylation of the enzymes acetyl-CoA carboxylase and hydroxymethylglutaryl-CoA reductase and mediates responses of these pathways to cellular stresses such as heat shock and depletion of glucose and ATP. AMPK is a heterotrimeric complex comprised of a catalytic alpha subunit and two non-catalytic beta and gamma subunits that are believed to regulate the activity of the alpha subunit. Subunits of AMPK have a much wider distribution in non-lipogenic tissues such as brain, heart, spleen, and lung than expected. This distribution suggests that its role may extend beyond regulation of lipid metabolism alone.

<u>Kinases in Apoptosis</u>

Apoptosis is a highly regulated signaling pathway leading to cell death that plays a crucial role in tissue development and homeostasis. Deregulation of this process is associated with the pathogenesis of a number of diseases including autoimmune diseases, neurodegenerative disorders, and cancer. Various STKs play key roles in this process. ZIP kinase is an STK containing a

C-terminal leucine zipper domain in addition to its N-terminal protein kinase domain. This C-terminal domain appears to mediate homodimerization and activation of the kinase as well as interactions with transcription factors such as activating transcription factor, ATF4, a member of the cyclic-AMP responsive element binding protein (ATF/CREB) family of transcriptional factors (Sanjo, H. et al. (1998) J. Biol. Chem. 273:29066-29071). DRAK1 and DRAK2 are STKs that share homology with the death-associated protein kinases (DAP kinases), known to function in interferon-γ induced apoptosis (Sanjo et al., *supra*). Like ZIP kinase, DAP kinases contain a C-terminal protein-protein interaction domain, in the form of ankyrin repeats, in addition to the N-terminal kinase domain. ZIP, DAP, and DRAK kinases induce morphological changes associated with apoptosis when transfected into NIH3T3 cells (Sanjo et al., *supra*). However, deletion of either the N-terminal kinase catalytic domain or the C-terminal domain of these proteins abolishes apoptosis activity, indicating that in addition to the kinase activity, activity in the C-terminal domain is also necessary for apoptosis, possibly as an interacting domain with a regulator or a specific substrate.

RICK is another STK recently identified as mediating a specific apoptotic pathway involving the death receptor, CD95 (Inohara, N. et al. (1998) J. Biol. Chem. 273:12296-12300). CD95 is a member of the tumor necrosis factor receptor superfamily and plays a critical role in the regulation and homeostasis of the immune system (Nagata, S. (1997) Cell 88:355-365). The CD95 receptor signaling pathway involves recruitment of various intracellular molecules to a receptor complex following ligand binding. This process includes recruitment of the cysteine protease caspase-8 which, in turn, activates a caspase cascade leading to cell death. RICK is composed of an N-terminal kinase catalytic domain and a C-terminal "caspase-recruitment" domain that interacts with caspase-like domains, indicating that RICK plays a role in the recruitment of caspase-8. This interpretation is supported by the fact that the expression of RICK in human 293T cells promotes activation of caspase-8 and potentiates the induction of apoptosis by various proteins involved in the CD95 apoptosis pathway (Inohara et al., *supra*).

<u>Mitochondrial Protein Kinases</u>

A novel class of eukaryotic kinases, related by sequence to prokaryotic histidine protein kinases, are the mitochondrial protein kinases (MPKs) which seem to have no sequence similarity with other eukaryotic protein kinases. These protein kinases are located exclusively in the mitochondrial matrix space and may have evolved from genes originally present in respiration-dependent bacteria which were endocytosed by primitive eukaryotic cells. MPKs are responsible for phosphorylation and inactivation of the branched-chain alpha-ketoacid dehydrogenase and pyruvate dehydrogenase complexes (Harris, R.A. et al. (1995) Adv. Enzyme Regul. 34:147-162). Five MPKs have been identified. Four members correspond to pyruvate dehydrogenase kinase isozymes, regulating the activity of the pyruvate dehydrogenase complex, which is an important regulatory

enzyme at the interface between glycolysis and the citric acid cycle. The fifth member corresponds to a branched-chain alpha-ketoacid dehydrogenase kinase, important in the regulation of the pathway for the disposal of branched-chain amino acids. (Harris, R.A. et al. (1997) Adv. Enzyme Regul. 37:271-293). Both starvation and the diabetic state are known to result in a great increase in the activity of the pyruvate dehydrogenase kinase in the liver, heart and muscle of the rat. This increase contributes in both disease states to the phosphorylation and inactivation of the pyruvate dehydrogenase complex and conservation of pyruvate and lactate for gluconeogenesis (Harris (1995) *supra*).

KINASES WITH NON-PROTEIN SUBSTRATES

Lipid and Inositol kinases

Lipid kinases phosphorylate hydroxyl residues on lipid head groups. A family of kinases involved in phosphorylation of phosphatidylinositol (PI) has been described, each member phosphorylating a specific carbon on the inositol ring (Leevers, S.J. et al. (1999) Curr. Opin. Cell. Biol. 11:219-225). The phosphorylation of phosphatidylinositol is involved in activation of the protein kinase C signaling pathway. The inositol phospholipids (phosphoinositides) intracellular signaling pathway begins with binding of a signaling molecule to a G-protein linked receptor in the plasma membrane. This leads to the phosphorylation of phosphatidylinositol (PI) residues on the inner side of the plasma membrane by inositol kinases, thus converting PI residues to the biphosphate state (PIP_2). PIP_2 is then cleaved into inositol triphosphate (IP_3) and diacylglycerol. These two products act as mediators for separate signaling pathways. Cellular responses that are mediated by these pathways are glycogen breakdown in the liver in response to vasopressin, smooth muscle contraction in response to acetylcholine, and thrombin-induced platelet aggregation.

PI 3-kinase (PI3K), which phosphorylates the D3 position of PI and its derivatives, has a central role in growth factor signal cascades involved in cell growth, differentiation, and metabolism. PI3K is a heterodimer consisting of an adapter subunit and a catalytic subunit. The adapter subunit acts as a scaffolding protein, interacting with specific tyrosine-phosphorylated proteins, lipid moieties, and other cytosolic factors. When the adapter subunit binds tyrosine phosphorylated targets, such as the insulin responsive substrate (IRS)-1, the catalytic subunit is activated and converts PI (4,5) bisphosphate (PIP_2) to PI (3,4,5) P_3 (PIP_3). PIP_3 then activates a number of other proteins, including PKA, protein kinase B (PKB), protein kinase C (PKC), glycogen synthase kinase (GSK)-3, and p70 ribosomal s6 kinase. PI3K also interacts directly with the cytoskeletal organizing proteins, Rac, rho, and cdc42 (Shepherd, P.R. et al. (1998) Biochem. J. 333:471-490). Animal models for diabetes, such as *obese* and *fat* mice, have altered PI3K adapter subunit levels. Specific mutations in the adapter subunit have also been found in an insulin-resistant Danish population, suggesting a role for PI3K in type-2 diabetes (Shepard, *supra*).

An example of lipid kinase phosphorylation activity is the phosphorylation of

D-erythro-sphingosine to the sphingolipid metabolite, sphingosine-1-phosphate (SPP). SPP has emerged as a novel lipid second-messenger with both extracellular and intracellular actions (Kohama, T. et al. (1998) J. Biol. Chem. 273:23722-23728). Extracellularly, SPP is a ligand for the G-protein coupled receptor EDG-1 (endothelial-derived, G-protein coupled receptor). Intracellularly, SPP regulates cell growth, survival, motility, and cytoskeletal changes. SPP levels are regulated by sphingosine kinases that specifically phosphorylate D-erythro-sphingosine to SPP. The importance of sphingosine kinase in cell signaling is indicated by the fact that various stimuli, including platelet-derived growth factor (PDGF), nerve growth factor, and activation of protein kinase C, increase cellular levels of SPP by activation of sphingosine kinase, and the fact that competitive inhibitors of the enzyme selectively inhibit cell proliferation induced by PDGF (Kohama et al., *supra*).

<u>Purine Nucleotide Kinases</u>

The purine nucleotide kinases, adenylate kinase (ATP:AMP phosphotransferase, or AdK) and guanylate kinase (ATP:GMP phosphotransferase, or GuK) play a key role in nucleotide metabolism and are crucial to the synthesis and regulation of cellular levels of ATP and GTP, respectively. These two molecules are precursors in DNA and RNA synthesis in growing cells and provide the primary source of biochemical energy in cells (ATP), and signal transduction pathways (GTP). Inhibition of various steps in the synthesis of these two molecules has been the basis of many antiproliferative drugs for cancer and antiviral therapy (Pillwein, K. et al. (1990) Cancer Res. 50:1576-1579).

AdK is found in almost all cell types and is especially abundant in cells having high rates of ATP synthesis and utilization such as skeletal muscle. In these cells AdK is physically associated with mitochondria and myofibrils, the subcellular structures that are involved in energy production and utilization, respectively. Recent studies have demonstrated a major function for AdK in transferring high energy phosphoryls from metabolic processes generating ATP to cellular components consuming ATP (Zeleznikar, R.J. et al. (1995) J. Biol. Chem. 270:7311-7319). Thus AdK may have a pivotal role in maintaining energy production in cells, particularly those having a high rate of growth or metabolism such as cancer cells, and may provide a target for suppression of its activity in order to treat certain cancers. Alternatively, reduced AdK activity may be a source of various metabolic, muscle-energy disorders that can result in cardiac or respiratory failure and may be treatable by increasing AdK activity.

GuK, in addition to providing a key step in the synthesis of GTP for RNA and DNA synthesis, also fulfills an essential function in signal transduction pathways of cells through the regulation of GDP and GTP. Specifically, GTP binding to membrane associated G proteins mediates the activation of cell receptors, subsequent intracellular activation of adenyl cyclase, and production of the second messenger, cyclic AMP. GDP binding to G proteins inhibits these processes. GDP and

GTP levels also control the activity of certain oncogenic proteins such as $p21^{ras}$ known to be involved in control of cell proliferation and oncogenesis (Bos, J.L. (1989) Cancer Res. 49:4682-4689). High ratios of GTP:GDP caused by suppression of GuK cause activation of $p21^{ras}$ and promote oncogenesis. Increasing GuK activity to increase levels of GDP and reduce the GTP:GDP ratio may provide a therapeutic strategy to reverse oncogenesis.

GuK is an important enzyme in the phosphorylation and activation of certain antiviral drugs useful in the treatment of herpes virus infections. These drugs include the guanine homologs acyclovir and buciclovir (Miller, W.H. and R.L. Miller (1980) J. Biol. Chem. 255:7204-7207; Stenberg, K. et al. (1986) J. Biol. Chem. 261:2134-2139). Increasing GuK activity in infected cells may provide a therapeutic strategy for augmenting the effectiveness of these drugs and possibly for reducing the necessary dosages of the drugs.

<u>Pyrimidine Kinases</u>

The pyrimidine kinases are deoxycytidine kinase and thymidine kinase 1 and 2. Deoxycytidine kinase is located in the nucleus, and thymidine kinase 1 and 2 are found in the cytosol (Johansson, M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:11941-11945). Phosphorylation of deoxyribonucleosides by pyrimidine kinases provides an alternative pathway for *de novo* synthesis of DNA precursors. The role of pyrimidine kinases, like purine kinases, in phosphorylation is critical to the activation of several chemotherapeutically important nucleoside analogues (Arner E.S. and S. Eriksson (1995) Pharmacol. Ther. 67:155-186).

<u>PHOSPHATASES</u>

Protein phosphatases are generally characterized as either serine/threonine- or tyrosine-specific based on their preferred phospho-amino acid substrate. However, some phosphatases (DSPs, for dual specificity phosphatases) can act on phosphorylated tyrosine, serine, or threonine residues. The protein serine/threonine phosphatases (PSPs) are important regulators of many cAMP-mediated hormone responses in cells. Protein tyrosine phosphatases (PTPs) play a significant role in cell cycle and cell signaling processes. Another family of phosphatases is the acid phosphatase or histidine acid phosphatase (HAP) family whose members hydrolyze phosphate esters at acidic pH conditions.

PSPs are found in the cytosol, nucleus, and mitochondria and in association with cytoskeletal and membranous structures in most tissues, especially the brain. Some PSPs require divalent cations, such as Ca^{2+} or Mn^{2+}, for activity. PSPs play important roles in glycogen metabolism, muscle contraction, protein synthesis, T cell function, neuronal activity, oocyte maturation, and hepatic metabolism (reviewed in Cohen, P. (1989) Annu. Rev. Biochem. 58:453-508). PSPs can be separated into two classes. The PPP class includes PP1, PP2A, PP2B/calcineurin, PP4, PP5, PP6, and PP7. Members of this class are composed of a homologous catalytic subunit bearing a very highly conserved signature sequence, coupled with one or more regulatory subunits (PROSITE

PDOC00115). Further interactions with scaffold and anchoring molecules determine the intracellular localization of PSPs and substrate specificity. The PPM class consists of several closely related isoforms of PP2C and is evolutionarily unrelated to the PPP class.

PP1 dephosphorylates many of the proteins phosphorylated by cyclic AMP-dependent protein kinase (PKA) and is an important regulator of many cAMP-mediated hormone responses in cells. A number of isoforms have been identified, with the alpha and beta forms being produced by alternative splicing of the same gene. Both ubiquitous and tissue-specific targeting proteins for PP1 have been identified. In the brain, inhibition of PP1 activity by the dopamine and adenosine 3',5'-monophosphate-regulated phosphoprotein of 32kDa (DARPP-32) is necessary for normal dopamine response in neostriatal neurons (reviewed in Price, N.E. and M.C. Mumby (1999) Curr. Opin. Neurobiol. 9:336-342). PP1, along with PP2A, has been shown to limit motility in microvascular endothelial cells, suggesting a role for PSPs in the inhibition of angiogenesis (Gabel, S. et al. (1999) Otolaryngol. Head Neck Surg.121:463-468).

PP2A is the main serine/threonine phosphatase. The core PP2A enzyme consists of a single 36 kDa catalytic subunit (C) associated with a 65 kDa scaffold subunit (A), whose role is to recruit additional regulatory subunits (B). Three gene families encoding B subunits are known (PR55, PR61, and PR72), each of which contain multiple isoforms, and additional families may exist (Millward, T.A et al. (1999) Trends Biosci. 24:186-191). These "B-type" subunits are cell type- and tissue-specific and determine the substrate specificity, enzymatic activity, and subcellular localization of the holoenzyme. The PR55 family is highly conserved and bears a conserved motif (PROSITE PDOC00785). PR55 increases PP2A activity toward mitogen-activated protein kinase (MAPK) and MAPK kinase (MEK). PP2A dephosphorylates the MAPK active site, inhibiting the cell's entry into mitosis. Several proteins can compete with PR55 for PP2A core enzyme binding, including the CKII kinase catalytic subunit, polyomavirus middle and small T antigens, and SV40 small t antigen. Viruses may use this mechanism to commandeer PP2A and stimulate progression of the cell through the cell cycle (Pallas, D.C. et al. (1992) J. Virol. 66:886-893). Altered MAP kinase expression is also implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development. PP2A, in fact, can dephosphorylate and modulate the activities of more than 30 protein kinases *in vitro*, and other evidence suggests that the same is true *in vivo* for such kinases as PKB, PKC, the calmodulin-dependent kinases, ERK family MAP kinases, cyclin-dependent kinases, and the IκB kinases (reviewed in Millward et al., *supra*). PP2A is itself a substrate for CKI and CKII kinases, and can be stimulated by polycationic macromolecules. A PP2A-like phosphatase is necessary to maintain the G1 phase destruction of mammalian cyclins A and B (Bastians, H. et al. (1999) Mol. Biol. Cell 10:3927-3941). PP2A is a major activity in the brain and is implicated in regulating neurofilament stability and normal neural function, particularly the

phosphorylation of the microtubule-associated protein tau. Hyperphosphorylation of tau has been proposed to lead to the neuronal degeneration seen in Alzheimer's disease (reviewed in Price and Mumby, *supra*).

PP2B, or calcineurin, is a Ca^{2+}-activated dimeric phosphatase and is particularly abundant in the brain. It consists of catalytic and regulatory subunits, and is activated by the binding of the calcium/calmodulin complex. Calcineurin is the target of the immunosuppressant drugs cyclosporine and FK506. Along with other cellular factors, these drugs interact with calcineurin and inhibit phosphatase activity. In T cells, this blocks the calcium dependent activation of the NF-AT family of transcription factors, leading to immunosuppression. This family is widely distributed, and it is likely that calcineurin regulates gene expression in other tissues as well. In neurons, calcineurin modulates functions which range from the inhibition of neurotransmitter release to desensitization of postsynaptic NMDA-receptor coupled calcium channels to long term memory (reviewed in Price and Mumby, *supra*).

Other members of the PPP class have recently been identified (Cohen, P.T. (1997) Trends Biochem. Sci. 22:245-251). One of them, PP5, contains regulatory domains with tetratricopeptide repeats. It can be activated by polyunsaturated fatty acids and anionic phospholipids *in vitro* and appears to be involved in a number of signaling pathways, including those controlled by atrial natriuretic peptide or steroid hormones (reviewed in Andreeva, A.V. and M.A. Kutuzov (1999) Cell Signal. 11:555-562).

PP2C is a ~42kDa monomer with broad substrate specificity and is dependent on divalent cations (mainly Mn^{2+} or Mg^{2+}) for its activity. PP2C proteins share a conserved N-terminal region with an invariant DGH motif, which contains an aspartate residue involved in cation binding (PROSITE PDOC00792). Targeting proteins and mechanisms regulating PP2C activity have not been identified. PP2C has been shown to inhibit the stress-responsive p38 and Jun kinase (JNK) pathways (Takekawa, M. et al. (1998) EMBO J. 17:4744-4752).

In contrast to PSPs, tyrosine-specific phosphatases (PTPs) are generally monomeric proteins of very diverse size (from 20kDa to greater than 100kDa) and structure that function primarily in the transduction of signals across the plasma membrane. PTPs are categorized as either soluble phosphatases or transmembrane receptor proteins that contain a phosphatase domain. All PTPs share a conserved catalytic domain of about 300 amino acids which contains the active site. The active site consensus sequence includes a cysteine residue which executes a nucleophilic attack on the phosphate moiety during catalysis (Neel, B.G. and N.K. Tonks (1997) Curr. Opin. Cell Biol. 9:193-204). Receptor PTPs are made up of an N-terminal extracellular domain of variable length, a transmembrane region, and a cytoplasmic region that generally contains two copies of the catalytic domain. Although only the first copy seems to have enzymatic activity, the second copy apparently

affects the substrate specificity of the first. The extracellular domains of some receptor PTPs contain fibronectin-like repeats, immunoglobulin-like domains, MAM domains (an extracellular motif likely to have an adhesive function), or carbonic anhydrase-like domains (PROSITE PDOC 00323). This wide variety of structural motifs accounts for the diversity in size and specificity of PTPs.

PTPs play important roles in biological processes such as cell adhesion, lymphocyte activation, and cell proliferation. PTPs μ and κ are involved in cell-cell contacts, perhaps regulating cadherin/catenin function. A number of PTPs affect cell spreading, focal adhesions, and cell motility, most of them via the integrin/tyrosine kinase signaling pathway (reviewed in Neel and Tonks, *supra*). CD45 phosphatases regulate signal transduction and lymphocyte activation (Ledbetter, J.A. et al. (1988) Proc. Natl. Acad. Sci. USA 85:8628-8632). Soluble PTPs containing Src-homology-2 domains have been identified (SHPs), suggesting that these molecules might interact with receptor tyrosine kinases. SHP-1 regulates cytokine receptor signaling by controlling the Janus family PTKs in hematopoietic cells, as well as signaling by the T-cell receptor and c-Kit (reviewed in Neel and Tonks, *supra*). M-phase inducer phosphatase plays a key role in the induction of mitosis by dephosphorylating and activating the PTK CDC2, leading to cell division (Sadhu, K. et al. (1990) Proc. Natl. Acad. Sci. USA 87:5139-5143). In addition, the genes encoding at least eight PTPs have been mapped to chromosomal regions that are translocated or rearranged in various neoplastic conditions, including lymphoma, small cell lung carcinoma, leukemia, adenocarcinoma, and neuroblastoma (reviewed in Charbonneau, H. and N.K. Tonks (1992) Annu. Rev. Cell Biol. 8:463-493). The PTP enzyme active site comprises the consensus sequence of the MTM1 gene family. The MTM1 gene is responsible for X-linked recessive myotubular myopathy, a congenital muscle disorder that has been linked to Xq28 (Kioschis, P. et al., (1998) Genomics 54:256-266). Many PTKs are encoded by oncogenes, and it is well known that oncogenesis is often accompanied by increased tyrosine phosphorylation activity. It is therefore possible that PTPs may serve to prevent or reverse cell transformation and the growth of various cancers by controlling the levels of tyrosine phosphorylation in cells. This is supported by studies showing that overexpression of PTP can suppress transformation in cells and that specific inhibition of PTP can enhance cell transformation (Charbonneau and Tonks, *supra*).

Dual specificity phosphatases (DSPs) are structurally more similar to the PTPs than the PSPs. DSPs bear an extended PTP active site motif with an additional 7 amino acid residues. DSPs are primarily associated with cell proliferation and include the cell cycle regulators cdc25A, B, and C. The phosphatases DUSP1 and DUSP2 inactivate the MAPK family members ERK (extracellular signal-regulated kinase), JNK (c-Jun N-terminal kinase), and p38 on both tyrosine and threonine residues (PROSITE PDOC 00323, *supra*). In the activated state, these kinases have been implicated in neuronal differentiation, proliferation, oncogenic transformation, platelet aggregation, and

apoptosis. Thus, DSPs are necessary for proper regulation of these processes (Muda, M. et al. (1996) J. Biol. Chem. 271:27205-27208). The tumor suppressor PTEN is a DSP that also shows lipid phosphatase activity. It seems to negatively regulate interactions with the extracellular matrix and maintains sensitivity to apoptosis. PTEN has been implicated in the prevention of angiogenesis (Giri, D. and M. Ittmann (1999) Hum. Pathol. 30:419-424) and abnormalities in its expression are associated with numerous cancers (reviewed in Tamura, M. et al. (1999) J. Natl. Cancer Inst. 91:1820-1828).

Histidine acid phosphatase (HAP; EXPASY EC 3.1.3.2), also known as acid phosphatase, hydrolyzes a wide spectrum of substrates including alkyl, aryl, and acyl orthophosphate monoesters and phosphorylated proteins at low pH. HAPs share two regions of conserved sequences, each centered around a histidine residue which is involved in catalytic activity. Members of the HAP family include lysosomal acid phosphatase (LAP) and prostatic acid phosphatase (PAP), both sensitive to inhibition by L-tartrate (PROSITE PDOC00538).

Synaptojanin, a polyphosphoinositide phosphatase, dephosphorylates phosphoinositides at positions 3, 4 and 5 of the inositol ring. Synaptojanin is a major presynaptic protein found at clathrin-coated endocytic intermediates in nerve terminals, and binds the clathrin coat-associated protein, EPS15. This binding is mediated by the C-terminal region of synaptojanin-170, which has 3 Asp-Pro-Phe amino acid repeats. Further, this 3 residue repeat had been found to be the binding site for the EH domains of EPS15 (Haffner, C. et al. (1997) FEBS Lett. 419:175-180). Additionally, synaptojanin may potentially regulate interactions of endocytic proteins with the plasma membrane, and be involved in synaptic vesicle recycling (Brodin, L. et al. (2000) Curr. Opin. Neurobiol. 10:312-320). Studies in mice with a targeted disruption in the synaptojanin 1 gene (Synj1) were shown to support coat formation of endocytic vesicles more effectively than was seen in wild-type mice, suggesting that Synj1 can act as a negative regulator of membrane-coat protein interactions. These findings provide genetic evidence for a crucial role of phosphoinositide metabolism in synaptic vesicle recycling (Cremona, O. et al. (1999) Cell 99:179-188).

Expression profiling

Microarrays are analytical tools used in bioanalysis. A microarray has a plurality of molecules spatially distributed over, and stably associated with, the surface of a solid support. Microarrays of polypeptides, polynucleotides, and/or antibodies have been developed and find use in a variety of applications, such as gene sequencing, monitoring gene expression, gene mapping, bacterial identification, drug discovery, and combinatorial chemistry.

One area in particular in which microarrays find use is in gene expression analysis. Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single

gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

Neurological disorders

Characterization of region-specific gene expression in the human brain provides a context and background for molecular neurobiology on a variety of neurological disorders. For example, Alzheimer's disease (AD) is a progressive, neurodestructive process of the human neocortex, characterized by the deterioration of memory and higher cognitive function. A progressive and irreversible brain disorder, AD is characterized by three major pathogenic episodes involving (a) an aberrant processing and deposition of beta-amyloid precursor protein (betaAPP) to form neurotoxic beta-amyloid (betaA) peptides and an aggregated insoluble polymer of betaA that forms the senile plaque, (b) the establishment of intraneuronal neuritic tau pathology yielding widespread deposits of agyrophilic neurofibrillary tangles (NFT) and (c) the initiation and proliferation of a brain-specific inflammatory response. These three seemingly disperse attributes of AD etiopathogenesis are linked by the fact that proinflammatory microglia, reactive astrocytes and their associated cytokines and chemokines are associated with the biology of the microtubule associated protein tau, betaA speciation and aggregation. Missense mutations in the presenilin genes PS1 and PS2, implicated in early onset familial AD, cause abnormal betaAPP processing with resultant overproduction of betaA42 and related neurotoxic peptides. Specific betaA fragments such as betaA42 can further potentiate proinflammatory mechanisms. Expression of the inducible oxidoreductase cyclooxygenase-2 and cytosolic phospholipase A2 (cPLA2) is strongly activated during cerebral ischemia and trauma, epilepsy and AD, indicating the induction of proinflammatory gene pathways as a response to brain injury. Neurotoxic metals such as aluminum and zinc, both implicated in AD etiopathogenesis, and arachidonic acid, a major metabolite of brain cPLA2 activity, each polymerize hyperphosphorylated tau to form NFT-like bundles. Studies have identified a reduced risk for AD in patients aged over 70 years who were previously treated with non-steroidal anti-inflammatory drugs for non-CNS afflictions that include arthritis. (For a review of the interrelationships between the mechanisms of PS1, PS2 and betaAPP gene expression, tau and betaA deposition and the induction, regulation and proliferation in AD of the neuroinflammatory response, see Lukiw, W.J, and Bazan, N.G. (2000) Neurochem. Res. 2000 25:1173-1184).

Breast cancer

More than 180,000 new cases of breast cancer are diagnosed each year, and the mortality rate for breast cancer approaches 10% of all deaths in females between the ages of 45-54 (Gish, K. (1999)

AWIS Magazine 28:7-10). However the survival rate based on early diagnosis of localized breast cancer is extremely high (97%), compared with the advanced stage of the disease in which the tumor has spread beyond the breast (22%). Current procedures for clinical breast examination are lacking in sensitivity and specificity, and efforts are underway to develop comprehensive gene expression profiles for breast cancer that may be used in conjunction with conventional screening methods to improve diagnosis and prognosis of this disease (Perou, C.M. et al. (2000) Nature 406:747-752).

Mutations in two genes, BRCA1 and BRCA2, are known to greatly predispose a woman to breast cancer and may be passed on from parents to children (Gish, *supra*). However, this type of hereditary breast cancer accounts for only about 5% to 9% of breast cancers, while the vast majority of breast cancer is due to non-inherited mutations that occur in breast epithelial cells.

The relationship between expression of epidermal growth factor (EGF) and its receptor, EGFR, to human mammary carcinoma has been particularly well studied. (See Khazaie, K. et al. (1993) Cancer and Metastasis Rev. 12:255-274, and references cited therein for a review of this area.) Overexpression of EGFR, particularly coupled with down-regulation of the estrogen receptor, is a marker of poor prognosis in breast cancer patients. In addition, EGFR expression in breast tumor metastases is frequently elevated relative to the primary tumor, suggesting that EGFR is involved in tumor progression and metastasis. This is supported by accumulating evidence that EGF has effects on cell functions related to metastatic potential, such as cell motility, chemotaxis, secretion and differentiation. Changes in expression of other members of the erbB receptor family, of which EGFR is one, have also been implicated in breast cancer. The abundance of erbB receptors, such as HER-2/neu, HER-3, and HER-4, and their ligands in breast cancer points to their functional importance in the pathogenesis of the disease, and may therefore provide targets for therapy of the disease (Bacus, S.S. et al. (1994) Am. J. Clin. Pathol. 102:S13-S24). Other known markers of breast cancer include a human secreted frizzled protein mRNA that is downregulated in breast tumors; the matrix G1a protein which is overexpressed in human breast carcinoma cells; Drg1 or RTP, a gene whose expression is diminished in colon, breast, and prostate tumors; maspin, a tumor suppressor gene downregulated in invasive breast carcinomas; and CaN19, a member of the S100 protein family, all of which are down-regulated in mammary carcinoma cells relative to normal mammary epithelial cells (Zhou, Z. et al. (1998) Int. J. Cancer 78:95-99; Chen, L. et al. (1990) Oncogene 5:1391-1395; Ulrix, W. et al (1999) FEBS Lett 455:23-26; Sager, R. et al. (1996) Curr. Top. Microbiol. Immunol. 213:51-64; and Lee, S.W. et al. (1992) Proc. Natl. Acad. Sci. USA 89:2504-2508).

Cell lines derived from human mammary epithelial cells at various stages of breast cancer provide a useful model to study the process of malignant transformation and tumor progression as it has been shown that these cell lines retain many of the properties of their parental tumors for lengthy culture periods (Wistuba, I.I. et al. (1998) Clin. Cancer Res. 4:2931-2938). Such a model is

particularly useful for comparing phenotypic and molecular characteristics of human mammary epithelial cells at various stages of malignant transformation.

There is a need in the art for new compositions, including nucleic acids and proteins, for the diagnosis, prevention, and treatment of cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers.

SUMMARY OF THE INVENTION

Various embodiments of the invention provide purified polypeptides, kinases and phosphatases, referred to collectively as 'KPP' and individually as 'KPP-1,' 'KPP-2,' 'KPP-3,' 'KPP-4,' 'KPP-5,' 'KPP-6,' 'KPP-7,' 'KPP-8,' 'KPP-9,' 'KPP-10,' 'KPP-11,' 'KPP-12,' 'KPP-13,' 'KPP-14,' and 'KPP-15' and methods for using these proteins and their encoding polynucleotides for the detection, diagnosis, and treatment of diseases and medical conditions. Embodiments also provide methods for utilizing the purified kinases and phosphatases and/or their encoding polynucleotides for facilitating the drug discovery process, including determination of efficacy, dosage, toxicity, and pharmacology. Related embodiments provide methods for utilizing the purified kinases and phosphatases and/or their encoding polynucleotides for investigating the pathogenesis of diseases and medical conditions.

An embodiment provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. Another embodiment provides an isolated polypeptide comprising an amino acid sequence of SEQ ID NO:1-15.

Still another embodiment provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. In another embodiment, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-15. In an alternative embodiment, the polynucleotide is selected from the group consisting of SEQ ID NO:16-30.

Still another embodiment provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. Another embodiment provides a cell transformed with the recombinant polynucleotide. Yet another embodiment provides a transgenic organism comprising the recombinant polynucleotide.

Another embodiment provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Yet another embodiment provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.

Still yet another embodiment provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA

equivalent of a)-d). In other embodiments, the polynucleotide can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex. In a related embodiment, the method can include detecting the amount of the hybridization complex. In still other embodiments, the probe can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Still yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof. In a related embodiment, the method can include detecting the amount of the amplified target polynucleotide or fragment thereof.

Another embodiment provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and a pharmaceutically acceptable excipient. In one

embodiment, the composition can comprise an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. Other embodiments provide a method of treating a disease or condition associated with decreased or abnormal expression of functional KPP, comprising administering to a patient in need of such treatment the composition.

Yet another embodiment provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. The method comprises a) contacting a sample comprising the polypeptide with a compound, and b) detecting agonist activity in the sample. Another embodiment provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment the composition.

Still yet another embodiment provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. The method comprises a) contacting a sample comprising the polypeptide with a compound, and b) detecting antagonist activity in the sample. Another embodiment provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with overexpression of functional KPP, comprising administering to a patient in need of such treatment the composition.

Another embodiment provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active

fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

Yet another embodiment provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

Still yet another embodiment provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, the method comprising a) contacting a sample comprising the target polynucleotide with a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

Another embodiment provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide

complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide can comprise a fragment of a polynucleotide selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for full length polynucleotide and polypeptide embodiments of the invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptide embodiments of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide embodiments, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide embodiments, along with selected fragments of the polynucleotides.

Table 5 shows representative cDNA libraries for polynucleotide embodiments.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze polynucleotides and polypeptides, along with applicable descriptions, references, and threshold parameters.

Table 8 shows single nucleotide polymorphisms found in polynucleotide sequences of the invention, along with allele frequencies in different human populations.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleic acids, and methods are described, it is understood that embodiments of the invention are not limited to the particular machines, instruments, materials, and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the invention.

As used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with various embodiments of the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"KPP" refers to the amino acid sequences of substantially purified KPP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of KPP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of KPP either by directly interacting with KPP or by acting on components of the biological pathway in which KPP participates.

An "allelic variant" is an alternative form of the gene encoding KPP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding KPP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as KPP or a polypeptide with at least one functional characteristic of KPP. Included within this definition are

polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding KPP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide encoding KPP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent KPP. Deliberate amino acid substitutions may be made on the basis of one or more similarities in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of KPP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" can refer to an oligopeptide, a peptide, a polypeptide, or a protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid. Amplification may be carried out using polymerase chain reaction (PCR) technologies or other nucleic acid amplification technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of KPP. Antagonists may include proteins such as antibodies, anticalins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of KPP either by directly interacting with KPP or by acting on components of the biological pathway in which KPP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind KPP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that

makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH_2), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker (Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13).

The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl. Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a polynucleotide having a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand,

and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic KPP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide" and a "composition comprising a given polypeptide" can refer to any composition containing the given polynucleotide or polypeptide. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotides encoding KPP or fragments of KPP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (Accelrys, Burlington MA) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala

His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of KPP or a polynucleotide encoding KPP which can be identical in sequence to, but shorter in length than, the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from about 5 to about 1000 contiguous nucleotides or amino acid residues. A

fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:16-30 can comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:16-30, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:16-30 can be employed in one or more embodiments of methods of the invention, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:16-30 from related polynucleotides. The precise length of a fragment of SEQ ID NO:16-30 and the region of SEQ ID NO:16-30 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-15 is encoded by a fragment of SEQ ID NO:16-30. A fragment of SEQ ID NO:1-15 can comprise a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-15. For example, a fragment of SEQ ID NO:1-15 can be used as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-15. The precise length of a fragment of SEQ ID NO:1-15 and the region of SEQ ID NO:1-15 to which the fragment corresponds can be determined based on the intended purpose for the fragment using one or more analytical methods described herein or otherwise known in the art.

A "full length" polynucleotide is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, alternatively, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of identical nucleotide matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using one or more computer algorithms or programs known in the art or described herein. For example, percent identity

can be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989; CABIOS 5:151-153) and in Higgins, D.G. et al. (1992; CABIOS 8:189-191). For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms which can be used is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of identical residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide. The phrases "percent similarity" and "% similarity," as applied to polypeptide sequences, refer to the percentage of residue matches, including identical residue matches and conservative substitutions, between at least two polypeptide sequences aligned using a standardized algorithm. In contrast, conservative substitutions are not included in the calculation of percent identity between polypeptide sequences.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least

150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY, ch. 9).

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as

formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acids by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid present in solution and another nucleic acid immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or polynucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of KPP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of KPP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, antibodies, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, antibody, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of KPP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of KPP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding

sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an KPP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of KPP.

"Probe" refers to nucleic acids encoding KPP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acids. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in, for example, Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY), Ausubel, F.M. et al. (1999; Short Protocols in Molecular Biology, 4th ed., John Wiley & Sons, New York NY), and Innis, M. et al. (1990; PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000

nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a nucleic acid that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook and Russell (*supra*). The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA molecule, is composed of the same linear sequence of nucleotides as the reference DNA molecule with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing KPP, nucleic acids encoding KPP, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably at least about 75% free, and most preferably at least about 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based

on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In another embodiment, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook and Russell (*supra*).

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotides that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of,

for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity or sequence similarity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity or sequence similarity over a certain defined length of one of the polypeptides.

THE INVENTION

Various embodiments of the invention include new human kinases and phosphatases (KPP), the polynucleotides encoding KPP, and the use of these compositions for the diagnosis, treatment, or prevention of cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide embodiments of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to polypeptide embodiments of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column

3 shows the number of amino acid residues in each polypeptide. Column 4 shows amino acid residues comprising signature sequences, domains, motifs, potential phosphorylation sites, and potential glycosylation sites. Column 5 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are kinases and phosphatases. For example, SEQ ID NO:6 is 93% identical, from residue E39 to residue I490, to human multifunctional calcium/calmodulin-dependent protein kinase II delta2 isoform (GenBank ID g4426595) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 9.0E-255, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:6 also has homology to calcium-calmodulin dependent protein kinase II delta, a member of the multifunctional CAMKII family involved in Ca2+ regulated processes, of which the alternative form delta 3 is specifically upregulated in the myocardium of patients with heart failure, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:6 also contains a protein kinase domain and a serine/threonine protein kinase catalytic domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM and SMART databases of conserved protein families/domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses, and BLAST analyses against the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID NO:6 is a calcium-calmodulin dependent protein kinase. The foregoing provides evidence that SEQ ID NO:6 is a calcium-calmodulin dependent protein kinase. SEQ ID NO:1-5 and SEQ ID NO:7-15 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-15 are described in Table 7.

As shown in Table 4, full length polynucleotide embodiments were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 lists fragments of the polynucleotides which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:16-30 or that distinguish between SEQ ID NO:16-30 and related polynucleotides. Column 3 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide embodiments. Columns 4 and 5 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 3 relative to their respective

full length sequences.

The identification numbers in Column 3 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 2944771F7 is the identification number of an Incyte cDNA sequence, and BRAITUT23 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 72678960V1). Alternatively, the identification numbers in column 3 may refer to GenBank cDNAs or ESTs (e.g., g3422499) which contributed to the assembly of the full length polynucleotides. In addition, the identification numbers in column 3 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the identification numbers in column 3 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the identification numbers in column 3 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL_*XXXXXX*_N_1_N_2_*YYYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column 3 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is the identification number of a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	**Type of analysis and/or examples of programs**

GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotides which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotides. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

Table 8 shows single nucleotide polymorphisms (SNPs) found in polynucleotide sequences of the invention, along with allele frequencies in different human populations. Columns 1 and 2 show the polynucleotide sequence identification number (SEQ ID NO:) and the corresponding Incyte project identification number (PID) for polynucleotides of the invention. Column 3 shows the Incyte identification number for the EST in which the SNP was detected (EST ID), and column 4 shows the identification number for the SNP (SNP ID). Column 5 shows the position within the EST sequence at which the SNP is located (EST SNP), and column 6 shows the position of the SNP within the full-length polynucleotide sequence (CB1 SNP). Column 7 shows the allele found in the EST sequence. Columns 8 and 9 show the two alleles found at the SNP site. Column 10 shows the amino acid encoded by the codon including the SNP site, based upon the allele found in the EST. Columns 11-14 show the frequency of allele 1 in four different human populations. An entry of n/d (not detected) indicates that the frequency of allele 1 in the population was too low to be detected, while n/a (not available) indicates that the allele frequency was not determined for the population.

The invention also encompasses KPP variants. Various embodiments of KPP variants can have at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% amino acid

sequence identity to the KPP amino acid sequence, and can contain at least one functional or structural characteristic of KPP.

Various embodiments also encompass polynucleotides which encode KPP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:16-30, which encodes KPP. The polynucleotide sequences of SEQ ID NO:16-30, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses variants of a polynucleotide encoding KPP. In particular, such a variant polynucleotide will have at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% polynucleotide sequence identity to a polynucleotide encoding KPP. A particular aspect of the invention encompasses a variant of a polynucleotide comprising a sequence selected from the group consisting of SEQ ID NO:16-30 which has at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:16-30. Any one of the polynucleotide variants described above can encode a polypeptide which contains at least one functional or structural characteristic of KPP.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide encoding KPP. A splice variant may have portions which have significant sequence identity to a polynucleotide encoding KPP, but will generally have a greater or lesser number of nucleotides due to additions or deletions of blocks of sequence arising from alternate splicing during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to a polynucleotide encoding KPP over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide encoding KPP. For example, a polynucleotide comprising a sequence of SEQ ID NO:19 and a polynucleotide comprising a sequence of SEQ ID NO:20 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:21 and a polynucleotide comprising a sequence of SEQ ID NO:22 are splice variants of each other; and a polynucleotide comprising a sequence of SEQ ID NO:23 and a polynucleotide comprising a sequence of SEQ ID NO:24 are splice variants of each other. Any one of the splice variants

described above can encode a polypeptide which contains at least one functional or structural characteristic of KPP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding KPP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring KPP, and all such variations are to be considered as being specifically disclosed.

Although polynucleotides which encode KPP and its variants are generally capable of hybridizing to polynucleotides encoding naturally occurring KPP under appropriately selected conditions of stringency, it may be advantageous to produce polynucleotides encoding KPP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding KPP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of polynucleotides which encode KPP and KPP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic polynucleotide may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a polynucleotide encoding KPP or any fragment thereof.

Embodiments of the invention can also include polynucleotides that are capable of hybridizing to the claimed polynucleotides, and, in particular, to those having the sequences shown in SEQ ID NO:16-30 and fragments thereof, under various conditions of stringency (Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511). Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Biosciences, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Invitrogen, Carlsbad CA). Preferably, sequence preparation is automated with machines

such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Amersham Biosciences), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art (Ausubel et al., *supra*, ch. 7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

The nucleic acids encoding KPP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art (Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (BD Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-

specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotides or fragments thereof which encode KPP may be cloned in recombinant DNA molecules that direct expression of KPP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other polynucleotides which encode substantially the same or a functionally equivalent polypeptides may be produced and used to express KPP.

The polynucleotides of the invention can be engineered using methods generally known in the art in order to alter KPP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of KPP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, polynucleotides encoding KPP may be synthesized, in whole or in part, using one or more chemical methods well known in the art (Caruthers, M.H. et al. (1980)

Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232). Alternatively, KPP itself or a fragment thereof may be synthesized using chemical methods known in the art. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques (Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J.Y. et al. (1995) Science 269:202-204). Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of KPP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography (Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421). The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing (Creighton, *supra*, pp. 28-53).

In order to express a biologically active KPP, the polynucleotides encoding KPP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotides encoding KPP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of polynucleotides encoding KPP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where a polynucleotide sequence encoding KPP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

Methods which are well known to those skilled in the art may be used to construct expression vectors containing polynucleotides encoding KPP and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination (Sambrook and Russell, *supra*, ch. 1-4, and 8; Ausubel et al., *supra*, ch. 1, 3, and 15).

A variety of expression vector/host systems may be utilized to contain and express

polynucleotides encoding KPP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems (Sambrook and Russell, *supra*; Ausubel et al., *supra*; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355). Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of polynucleotides to the targeted organ, tissue, or cell population (Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5:350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90:6340-6344; Buller, R.M. et al. (1985) Nature 317:813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31:219-226; Verma, I.M. and N. Somia (1997) Nature 389:239-242). The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotides encoding KPP. For example, routine cloning, subcloning, and propagation of polynucleotides encoding KPP can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Invitrogen). Ligation of polynucleotides encoding KPP into the vector's multiple cloning site disrupts the *lac*Z gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509). When large quantities of KPP are needed, e.g. for the production of antibodies, vectors which direct high level expression of KPP may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of KPP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign polynucleotide sequences into the host genome for stable propagation (Ausubel et al., *supra*; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994)

Bio/Technology 12:181-184).

Plant systems may also be used for expression of KPP. Transcription of polynucleotides encoding KPP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection (The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196).

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, polynucleotides encoding KPP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses KPP in host cells (Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355).

For long term production of recombinant proteins in mammalian systems, stable expression of KPP in cell lines is preferred. For example, polynucleotides encoding KPP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *apr*⁻ cells, respectively (Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823). Also, antimetabolite, antibiotic, or herbicide

resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14). Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051). Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; BD Clontech), β-glucuronidase and its substrate β-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding KPP is inserted within a marker gene sequence, transformed cells containing polynucleotides encoding KPP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding KPP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the polynucleotide encoding KPP and that express KPP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of KPP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on KPP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art (Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding KPP include oligolabeling,

nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, polynucleotides encoding KPP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Biosciences, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with polynucleotides encoding KPP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode KPP may be designed to contain signal sequences which direct secretion of KPP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted polynucleotides or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant polynucleotides encoding KPP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric KPP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of KPP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion

proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the KPP encoding sequence and the heterologous protein sequence, so that KPP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In another embodiment, synthesis of radiolabeled KPP may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

KPP, fragments of KPP, or variants of KPP may be used to screen for compounds that specifically bind to KPP. One or more test compounds may be screened for specific binding to KPP. In various embodiments, 1, 2, 3, 4, 5, 10, 20, 50, 100, or 200 test compounds can be screened for specific binding to KPP. Examples of test compounds can include antibodies, anticalins, oligonucleotides, proteins (e.g., ligands or receptors), or small molecules.

In related embodiments, variants of KPP can be used to screen for binding of test compounds, such as antibodies, to KPP, a variant of KPP, or a combination of KPP and/or one or more variants KPP. In an embodiment, a variant of KPP can be used to screen for compounds that bind to a variant of KPP, but not to KPP having the exact sequence of a sequence of SEQ ID NO:1-15. KPP variants used to perform such screening can have a range of about 50% to about 99% sequence identity to KPP, with various embodiments having 60%, 70%, 75%, 80%, 85%, 90%, and 95% sequence identity.

In an embodiment, a compound identified in a screen for specific binding to KPP can be closely related to the natural ligand of KPP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner (Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):Chapter 5). In another embodiment, the compound thus identified can be a natural ligand of a receptor KPP (Howard, A.D. et al. (2001) Trends Pharmacol. Sci.22:132-140; Wise, A. et al. (2002) Drug Discovery Today 7:235-246).

In other embodiments, a compound identified in a screen for specific binding to KPP can be closely related to the natural receptor to which KPP binds, at least a fragment of the receptor, or a fragment of the receptor including all or a portion of the ligand binding site or binding pocket. For example, the compound may be a receptor for KPP which is capable of propagating a signal, or a decoy receptor for KPP which is not capable of propagating a signal (Ashkenazi, A. and V.M. Divit (1999) Curr. Opin. Cell Biol. 11:255-260; Mantovani, A. et al. (2001) Trends Immunol. 22:328-336).

The compound can be rationally designed using known techniques. Examples of such techniques include those used to construct the compound etanercept (ENBREL; Amgen Inc., Thousand Oaks CA), which is efficacious for treating rheumatoid arthritis in humans. Etanercept is an engineered p75 tumor necrosis factor (TNF) receptor dimer linked to the Fc portion of human IgG_1 (Taylor, P.C. et al. (2001) Curr. Opin. Immunol. 13:611-616).

In one embodiment, two or more antibodies having similar or, alternatively, different specificities can be screened for specific binding to KPP, fragments of KPP, or variants of KPP. The binding specificity of the antibodies thus screened can thereby be selected to identify particular fragments or variants of KPP. In one embodiment, an antibody can be selected such that its binding specificity allows for preferential identification of specific fragments or variants of KPP. In another embodiment, an antibody can be selected such that its binding specificity allows for preferential diagnosis of a specific disease or condition having increased, decreased, or otherwise abnormal production of KPP.

In an embodiment, anticalins can be screened for specific binding to KPP, fragments of KPP, or variants of KPP. Anticalins are ligand-binding proteins that have been constructed based on a lipocalin scaffold (Weiss, G.A. and H.B. Lowman (2000) Chem. Biol. 7:R177-R184; Skerra, A. (2001) J. Biotechnol. 74:257-275). The protein architecture of lipocalins can include a beta-barrel having eight antiparallel beta-strands, which supports four loops at its open end. These loops form the natural ligand-binding site of the lipocalins, a site which can be re-engineered *in vitro* by amino acid substitutions to impart novel binding specificities. The amino acid substitutions can be made using methods known in the art or described herein, and can include conservative substitutions (e.g., substitutions that do not alter binding specificity) or substitutions that modestly, moderately, or significantly alter binding specificity.

In one embodiment, screening for compounds which specifically bind to, stimulate, or inhibit KPP involves producing appropriate cells which express KPP, either as a secreted protein or on the cell membrane. Preferred cells can include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing KPP or cell membrane fractions which contain KPP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either KPP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with KPP, either in solution or affixed to a solid support, and detecting the binding of KPP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

An assay can be used to assess the ability of a compound to bind to its natural ligand and/or to inhibit the binding of its natural ligand to its natural receptors. Examples of such assays include radio-labeling assays such as those described in U.S. Patent No. 5,914,236 and U.S. Patent No. 6,372,724. In a related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a receptor) to improve or alter its ability to bind to its natural ligands (Matthews, D.J. and J.A. Wells. (1994) Chem. Biol. 1:25-30). In another related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a ligand) to improve or alter its ability to bind to its natural receptors (Cunningham, B.C. and J.A. Wells (1991) Proc. Natl. Acad. Sci. USA 88:3407-3411; Lowman, H.B. et al. (1991) J. Biol. Chem. 266:10982-10988).

KPP, fragments of KPP, or variants of KPP may be used to screen for compounds that modulate the activity of KPP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for KPP activity, wherein KPP is combined with at least one test compound, and the activity of KPP in the presence of a test compound is compared with the activity of KPP in the absence of the test compound. A change in the activity of KPP in the presence of the test compound is indicative of a compound that modulates the activity of KPP. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising KPP under conditions suitable for KPP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of KPP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding KPP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease (see, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337). For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding KPP may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding KPP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding KPP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress KPP, e.g., by secreting KPP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of KPP and kinases and phosphatases. In addition, examples of tissues expressing KPP can be found in Table 6 and can also be found in Example XI. Therefore, KPP appears to play a role in cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers. In the treatment of disorders associated with increased KPP expression or activity, it is desirable to decrease the expression or activity of KPP. In the treatment of disorders associated with decreased KPP expression or activity, it is desirable to increase the expression or activity of KPP.

Therefore, in one embodiment, KPP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP. Examples of such disorders include, but are not limited to, a cardiovascular disease such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary

hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; an immune system disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis,

inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a disorder affecting growth and development such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM_2 gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatoses, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, uterus, leukemias such as multiple myeloma, and lymphomas such as Hodgkin's disease.

In another embodiment, a vector capable of expressing KPP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified KPP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of KPP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP including, but not limited to, those listed above.

In a further embodiment, an antagonist of KPP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of KPP. Examples of such disorders include, but are not limited to, those cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers described above. In one aspect, an antibody which specifically binds KPP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express KPP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding KPP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of KPP including, but not limited to, those described above.

In other embodiments, any protein, agonist, antagonist, antibody, complementary sequence, or vector embodiments may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of KPP may be produced using methods which are generally known in the art. In particular, purified KPP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind KPP. Antibodies to KPP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. In an embodiment, neutralizing antibodies (i.e., those which inhibit dimer formation) can be used therapeutically. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have application in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with KPP or with any

fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to KPP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are substantially identical to a portion of the amino acid sequence of the natural protein. Short stretches of KPP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to KPP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce KPP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137).

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments which contain specific binding sites for KPP may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and

easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989) Science 246:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between KPP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering KPP epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for KPP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of KPP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple KPP epitopes, represents the average affinity, or avidity, of the antibodies for KPP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular KPP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the KPP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of KPP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of KPP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available (Catty, *supra*; Coligan et al., *supra*).

In another embodiment of the invention, polynucleotides encoding KPP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding KPP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments

can be designed from various locations along the coding or control regions of sequences encoding KPP (Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press, Totawa NJ).

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein (Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102:469-475; Scanlon, K.J. et al. (1995) FASEB J. 9:1288-1296). Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors (Miller, A.D. (1990) Blood 76:271-278; Ausubel et al., *supra*; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63:323-347). Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art (Rossi, J.J. (1995) Br. Med. Bull. 51:217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87:1308-1315; Morris, M.C. et al. (1997) Nucleic Acids Res. 25:2730-2736).

In another embodiment of the invention, polynucleotides encoding KPP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in KPP expression or regulation causes disease, the expression of KPP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in KPP are treated by constructing mammalian expression vectors encoding KPP and introducing these

vectors by mechanical means into KPP-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J.-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of KPP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (BD Clontech, Palo Alto CA). KPP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, *supra*)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding KPP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to KPP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding KPP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for

receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4$^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In an embodiment, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding KPP to cells which have one or more genetic abnormalities with respect to the expression of KPP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999; Annu. Rev. Nutr. 19:511-544) and Verma, I.M. and N. Somia (1997; Nature 18:389:239-242).

In another embodiment, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding KPP to target cells which have one or more genetic abnormalities with respect to the expression of KPP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing KPP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999; J. Virol. 73:519-532) and Xu, H. et al. (1994; Dev. Biol. 163:152-161). The manipulation of cloned herpesvirus sequences, the generation of

recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another embodiment, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding KPP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for KPP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of KPP-coding RNAs and the synthesis of high levels of KPP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of KPP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of RNA molecules encoding KPP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes may be prepared by any method known in the art for the synthesis of nucleic acid molecules.. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA molecules encoding KPP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous endonucleases.

In other embodiments of the invention, the expression of one or more selected polynucleotides of the present invention can be altered, inhibited, decreased, or silenced using RNA interference (RNAi) or post-transcriptional gene silencing (PTGS) methods known in the art. RNAi is a post-transcriptional mode of gene silencing in which double-stranded RNA (dsRNA) introduced into a targeted cell specifically suppresses the expression of the homologous gene (i.e., the gene bearing the sequence complementary to the dsRNA). This effectively knocks out or substantially reduces the expression of the targeted gene. PTGS can also be accomplished by use of DNA or DNA fragments as well. RNAi methods are described by Fire, A. et al. (1998; Nature 391:806-811) and Gura, T. (2000; Nature 404:804-808). PTGS can also be initiated by introduction of a complementary segment of DNA into the selected tissue using gene delivery and/or viral vector delivery methods described herein or known in the art.

RNAi can be induced in mammalian cells by the use of small interfering RNA also known as siRNA. siRNA are shorter segments of dsRNA (typically about 21 to 23 nucleotides in length) that result *in vivo* from cleavage of introduced dsRNA by the action of an endogenous ribonuclease.

siRNA appear to be the mediators of the RNAi effect in mammals. The most effective siRNAs appear to be 21 nucleotide dsRNAs with 2 nucleotide 3' overhangs. The use of siRNA for inducing RNAi in mammalian cells is described by Elbashir, S.M. et al. (2001; Nature 411:494-498).

siRNA can be generated indirectly by introduction of dsRNA into the targeted cell. Alternatively, siRNA can be synthesized directly and introduced into a cell by transfection methods and agents described herein or known in the art (such as liposome-mediated transfection, viral vector methods, or other polynucleotide delivery/introductory methods). Suitable siRNAs can be selected by examining a transcript of the target polynucleotide (e.g., mRNA) for nucleotide sequences downstream from the AUG start codon and recording the occurrence of each nucleotide and the 3' adjacent 19 to 23 nucleotides as potential siRNA target sites, with sequences having a 21 nucleotide length being preferred. Regions to be avoided for target siRNA sites include the 5' and 3' untranslated regions (UTRs) and regions near the start codon (within 75 bases), as these may be richer in regulatory protein binding sites. UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP endonuclease complex. The selected target sites for siRNA can then be compared to the appropriate genome database (e.g., human, etc.) using BLAST or other sequence comparison algorithms known in the art. Target sequences with significant homology to other coding sequences can be eliminated from consideration. The selected siRNAs can be produced by chemical synthesis methods known in the art or by *in vitro* transcription using commercially available methods and kits such as the SILENCER siRNA construction kit (Ambion, Austin TX).

In alternative embodiments, long-term gene silencing and/or RNAi effects can be induced in selected tissue using expression vectors that continuously express siRNA. This can be accomplished using expression vectors that are engineered to express hairpin RNAs (shRNAs) using methods known in the art (see, e.g., Brummelkamp, T.R. et al. (2002) Science 296:550-553; and Paddison, P.J. et al. (2002) Genes Dev. 16:948-958). In these and related embodiments, shRNAs can be delivered to target cells using expression vectors known in the art. An example of a suitable expression vector for delivery of siRNA is the PSILENCER1.0-U6 (circular) plasmid (Ambion). Once delivered to the target tissue, shRNAs are processed *in vivo* into siRNA-like molecules capable of carrying out gene-specific silencing.

In various embodiments, the expression levels of genes targeted by RNAi or PTGS methods can be determined by assays for mRNA and/or protein analysis. Expression levels of the mRNA of a targeted gene can be determined, for example, by northern analysis methods using the NORTHERNMAX-GLY kit (Ambion); by microarray methods; by PCR methods; by real time PCR methods; and by other RNA/polynucleotide assays known in the art or described herein. Expression levels of the protein encoded by the targeted gene can be determined, for example, by microarray methods; by polyacrylamide gel electrophoresis; and by Western analysis using standard techniques

known in the art.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding KPP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased KPP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding KPP may be therapeutically useful, and in the treatment of disorders associated with decreased KPP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding KPP may be therapeutically useful.

In various embodiments, one or more test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding KPP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding KPP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding KPP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified

oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art (Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466).

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of KPP, antibodies to KPP, and mimetics, agonists, antagonists, or inhibitors of KPP.

In various embodiments, the compositions described herein, such as pharmaceutical compositions, may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery allows administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising KPP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the

macromolecule. Alternatively, KPP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example KPP or fragments thereof, antibodies of KPP, and agonists, antagonists or inhibitors of KPP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind KPP may be used for the diagnosis of disorders characterized by expression of KPP, or in assays to monitor patients being treated with KPP or agonists, antagonists, or inhibitors of KPP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for KPP include methods which utilize the antibody and a label to detect KPP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring KPP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of KPP expression. Normal or standard values for KPP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to KPP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of KPP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, polynucleotides encoding KPP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotides, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of KPP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of KPP, and to monitor regulation of KPP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotides, including genomic sequences, encoding KPP or closely related molecules may be used to identify nucleic acid sequences which encode KPP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding KPP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the KPP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:16-30 or from genomic sequences including promoters, enhancers, and introns of the KPP gene.

Means for producing specific hybridization probes for polynucleotides encoding KPP include the cloning of polynucleotides encoding KPP or KPP derivatives into vectors for the production of

mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotides encoding KPP may be used for the diagnosis of disorders associated with expression of KPP. Examples of such disorders include, but are not limited to, a cardiovascular disease such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; an immune system disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or

pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a disorder affecting growth and development such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy,

adrenoleukodystrophy, GM_2 gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatoses, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, uterus, leukemias such as multiple myeloma, and lymphomas such as Hodgkin's disease. Polynucleotides encoding KPP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered KPP expression. Such qualitative or quantitative methods are well known in the art.

In a particular embodiment, polynucleotides encoding KPP may be used in assays that detect the presence of associated disorders, particularly those mentioned above. Polynucleotides complementary to sequences encoding KPP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of polynucleotides encoding KPP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of KPP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding KPP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used.

Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier, thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding KPP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding KPP, or a fragment of a polynucleotide complementary to the polynucleotide encoding KPP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from polynucleotides encoding KPP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from polynucleotides encoding KPP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory

preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641).

Methods which may also be used to quantify the expression of KPP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves (Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotides described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the

fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, KPP, fragments of KPP, or antibodies specific for KPP may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time (Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484; hereby expressly incorporated by reference herein). Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*, as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity (see, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at

niehs.nih.gov/oc/news/toxchip.htm). Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In an embodiment, the toxicity of a test compound can be assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another embodiment relates to the use of the polypeptides disclosed herein to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of interest. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for KPP to quantify the levels of KPP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by contacting the microarray with the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol-

or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art (Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/25116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662). Various types of microarrays are well known and thoroughly described in Schena, M., ed. (1999; DNA Microarrays: A Practical Approach, Oxford University Press, London).

In another embodiment of the invention, nucleic acid sequences encoding KPP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal

mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; Trask, B.J. (1991) Trends Genet. 7:149-154). Once mapped, the nucleic acid sequences may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP) (Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357).

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data (Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968). Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding KPP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation (Gatti, R.A. et al. (1988) Nature 336:577-580). The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, KPP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between KPP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest (Geysen, et al. (1984) PCT application WO84/03564). In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with KPP, or fragments thereof, and washed.

Bound KPP is then detected by methods well known in the art. Purified KPP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding KPP specifically compete with a test compound for binding KPP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with KPP.

In additional embodiments, the nucleotide sequences which encode KPP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, are expressly incorporated by reference herein.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs are derived from cDNA libraries described in the LIFESEQ database (Incyte, Palo Alto CA) and shown in Table 4, column 3. Some tissues are homogenized and lysed in guanidinium isothiocyanate, while others are homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Invitrogen), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates are centrifuged over CsCl cushions or extracted with chloroform. RNA is precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA are repeated as necessary to increase RNA purity. In some cases, RNA is treated with DNase. For most libraries, poly(A)+ RNA is isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA is isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene is provided with RNA and constructs the corresponding cDNA libraries. Otherwise, cDNA is synthesized and cDNA libraries are constructed with the UNIZAP

vector system (Stratagene) or SUPERSCRIPT plasmid system (Invitrogen), using the recommended procedures or similar methods known in the art (Ausubel et al., *supra*, ch. 5). Reverse transcription is initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters are ligated to double stranded cDNA, and the cDNA is digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA is size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Biosciences) or preparative agarose gel electrophoresis. cDNAs are ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Invitrogen, Carlsbad CA), PCDNA2.1 plasmid (Invitrogen), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte, Palo Alto CA), pRARE (Incyte), or pINCY (Incyte), or derivatives thereof. Recombinant plasmids are transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Invitrogen.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I are recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids are purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids are resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA is amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps are carried out in a single reaction mixture. Samples are processed and stored in 384-well plates, and the concentration of amplified plasmid DNA is quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II are sequenced as follows. Sequencing reactions are processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions are prepared using reagents provided by Amersham Biosciences or supplied in ABI sequencing kits such as the ABI

PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides are carried out using the MEGABACE 1000 DNA sequencing system (Amersham Biosciences); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences are identified using standard methods (Ausubel et al., *supra*, ch. 7). Some of the cDNA sequences are selected for extension using the techniques disclosed in Example VIII.

Polynucleotide sequences derived from Incyte cDNAs are validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof are then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans* (Incyte, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz, J. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families; see, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries are performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences are assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) are used to extend Incyte cDNA assemblages to full length. Assembly is performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages are screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences are translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences are subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (MiraiBio, Alameda CA) and

LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences are also used to identify polynucleotide sequence fragments from SEQ ID NO:16-30. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative kinases and phosphatases are initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94; Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once is set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode kinases and phosphatases, the encoded polypeptides are analyzed by querying against PFAM models for kinases and phosphatases. Potential kinases and phosphatases are also identified by homology to Incyte cDNA sequences that have been annotated as kinases and phosphatases. These selected Genscan-predicted sequences are then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences are then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis is also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage is available, this information is used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences are obtained by assembling

Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences are derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences are extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III are mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster is analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that are subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval is present on more than one sequence in the cluster are identified, and intervals thus identified are considered to be equivalent by transitivity. For example, if an interval is present on a cDNA and two genomic sequences, then all three intervals are considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified are then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) are given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences are translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan are corrected by comparison to the top BLAST hit from genpept. Sequences are further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences are extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III are queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog is then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein is generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both are used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences are therefore

"stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences are examined to determine whether they contain a complete gene.

VI. Chromosomal Mapping of KPP Encoding Polynucleotides

The sequences used to assemble SEQ ID NO:16-30 are compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:16-30 are assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon are used to determine if any of the clustered sequences have been previously mapped. Inclusion of a mapped sequence in a cluster results in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook and Russell, *supra*, ch. 7; Ausubel et al., *supra*, ch. 4).

Analogous computer techniques applying BLAST are used to search for identical or related molecules in databases such as GenBank or LIFESEQ (Incyte). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{5 \text{ x minimum } \{\text{length(Seq. 1), length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotides encoding KPP are analyzed with respect to the tissue sources from which they are derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding KPP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ database (Incyte, Palo Alto CA).

VIII. Extension of KPP Encoding Polynucleotides

Full length polynucleotides are produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer is synthesized to initiate 5' extension of the known fragment, and the other primer is synthesized to initiate 3'

extension of the known fragment. The initial primers are designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations is avoided.

Selected human cDNA libraries are used to extend the sequence. If more than one extension is necessary or desired, additional or nested sets of primers are designed.

High fidelity amplification is obtained by PCR using methods well known in the art. PCR is performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contains DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ are as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well is determined by dispensing 100 μl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate is scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture is analyzed by electrophoresis on a 1 % agarose gel to determine which reactions are successful in extending the sequence.

The extended nucleotides are desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). For shotgun sequencing, the digested nucleotides are separated on low concentration (0.6 to 0.8%) agarose gels, fragments are excised, and agar digested with Agar ACE (Promega). Extended clones are religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Biosciences), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells are selected on antibiotic-containing media, and individual colonies are picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells are lysed, and DNA is amplified by PCR using Taq DNA polymerase (Amersham

Biosciences) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA is quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries are reamplified using the same conditions as described above. Samples are diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Biosciences) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotides are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in KPP Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) are identified in SEQ ID NO:16-30 using the LIFESEQ database (Incyte). Sequences from the same gene are clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters is used to distinguish SNPs from other sequence variants. Preliminary filters remove the majority of basecall errors by requiring a minimum Phred quality score of 15, and remove sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis is applied to the original chromatogram files in the vicinity of the putative SNP. Clone error filters use statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters use statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removes duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs are selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprises 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezualan, and two Amish individuals. The African population comprises 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprises 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprises 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies are first analyzed in the Caucasian population; in some cases those SNPs which show no

allelic variance in this population are not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:16-30 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Biosciences), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Biosciences). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to NYTRAN PLUS nylon membranes (Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing; see, e.g., Baldeschweiler et al., *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena, M., ed. (1999) DNA Microarrays: A Practical Approach, Oxford University Press, London). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements (Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31).

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The

array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (BD Clontech, Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma-Aldrich, St. Louis MO) in 95% ethanol. Coated slides

are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a

cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte). Array elements that exhibit at least about a two-fold change in expression, a signal-to-background ratio of at least about 2.5, and an element spot size of at least about 40%, are considered to be differentially expressed.

Expression

For example, expression of SEQ ID NO:18 was downregulated in brain tissue affected by Alzheimer's Disease versus normal brain tissue as determined by microarray analysis. Specific dissected brain regions from the brain patients with AD were compared to dissected regions from normal brain. The diagnosis of normal or AD was established by a certified neuropathologist based on microscopic examination of multiple sections throughout the brain. Expression of SEQ ID NO:18 was decreased at least two-fold in 7 of 10 AD-affected tissue samples. Therefore, in various embodiments, SEQ ID NO:18 can be used for one or more of the following: i) monitoring treatment of Alzheimer's Disease, ii) diagnostic assays for Alzheimer's Disease, and iii) developing therapeutics and/or other treatments for Alzheimer's Disease as determined by microarray analysis.

As another example, SEQ ID NO:16 and SEQ ID NO:18 were downregulated in breast cancer cells versus nonmalignant mammary epithelial cells, as determined by microarray analysis. Cell lines compared included: a) MCF-10A, a breast mammary gland (luminal ductal characteristics)

cell line isolated from a 36-year-old woman with fibrocystic breast disease, b) MCF7, a nonmalignant breast adenocarcinoma cell line isolated from the pleural effusion of a 69-year-old female, c) BT-20, a breast carcinoma cell line derived *in vitro* from the cells emigrating out of thin slices of tumor mass isolated from a 74-year-old female, d) T-47D, a breast carcinoma cell line isolated from a pleural effusion obtained from a 54-year-old female with an infiltrating ductal carcinoma of the breast, e) Sk-BR-3, a breast adenocarcinoma cell line isolated from a malignant pleural effusion of a 43-year-old female, f) MDA-mb-231, a breast tumor cell line isolated from the pleural effusion of a 51-year-old female, g) MDA-mb-435S, a spindle-shaped strain that evolved from the parent line (435) isolated by R. Cailleau from pleural effusion of a 31-year-old female with metastatic, ductal adenocarcinoma of the breast, and h) HMEC, a primary breast epithelial cell line isolated from a normal donor. Expression of SEQ ID NO:16 was decreased at least two-fold in the Sk-BR-3, BT-20, MDA-mb-435S, T-47D, and MCF7 cell lines as compared to the normal breast epithelial cells. Expression of SEQ ID NO:18 was decreased at least two-fold in the MCF-10A, T-47D, Sk-BR-3, and MCF7 cell lines as compared to the normal breast epithelial cells. Therefore, in various embodiments, SEQ ID NO:16 and SEQ ID NO:18 can be used for one or more of the following: i) monitoring treatment of breast cancer, ii) diagnostic assays for breast cancer, and iii) developing therapeutics and/or other treatments for breast cancer as determined by microarray analysis.

As another example, SEQ ID NO:18 and SEQ ID NO:21 showed tissue-specific expression as determined by microarray analysis. RNA samples isolated from a variety of normal human tissues were compared to a common reference sample. Tissues contributing to the reference sample were selected for their ability to provide a complete distribution of RNA in the human body and include brain (4%), heart (7%), kidney (3%), lung (8%), placenta (46%), small intestine (9%), spleen (3%), stomach (6%), testis (9%), and uterus (5%). The normal tissues assayed were obtained from at least three different donors. RNA from each donor was separately isolated and individually hybridized to the microarray. Since these hybridization experiments were conducted using a common reference sample, differential expression values are directly comparable from one tissue to another. The expression of SEQ ID NO:18 was increased by at least two-fold in brain cortex tissue as compared to the reference sample. Therefore, SEQ ID NO:18 can be used as a tissue marker for brain cortex tissue. The expression of SEQ ID NO:21 was increased by at least two-fold in heart tissue as compared to the reference sample. Therefore, SEQ ID NO:21 can be used as a tissue marker for heart tissue.

XII. Complementary Polynucleotides

Sequences complementary to the KPP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring KPP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with

smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of KPP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the KPP-encoding transcript.

XIII. Expression of KPP

Expression and purification of KPP is achieved using bacterial or virus-based expression systems. For expression of KPP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express KPP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of KPP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding KPP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus (Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945).

In most expression systems, KPP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Biosciences). Following purification, the GST moiety can be proteolytically cleaved from KPP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). Purified KPP obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, XIX, XX, and XXI, where applicable.

XIV. Functional Assays

KPP function is assessed by expressing the sequences encoding KPP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV'SPORT plasmid (Invitrogen, Carlsbad CA) and PCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; BD Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994; Flow Cytometry, Oxford, New York NY).

The influence of KPP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding KPP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding KPP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of KPP Specific Antibodies

KPP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the KPP amino acid sequence is analyzed using LASERGENE software

(DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art (Ausubel et al., *supra*, ch. 11).

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity (Ausubel et al., *supra*). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-KPP activity by, for example, binding the peptide or KPP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI. Purification of Naturally Occurring KPP Using Specific Antibodies

Naturally occurring or recombinant KPP is substantially purified by immunoaffinity chromatography using antibodies specific for KPP. An immunoaffinity column is constructed by covalently coupling anti-KPP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Biosciences). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing KPP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of KPP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/KPP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and KPP is collected.

XVII. Identification of Molecules Which Interact with KPP

KPP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent (Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled KPP, washed, and any wells with labeled KPP complex are assayed. Data obtained using different concentrations of KPP are used to calculate values for the number, affinity, and association of KPP with the candidate molecules.

Alternatively, molecules interacting with KPP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989; Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (BD Clontech).

KPP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S.

Patent No. 6,057,101).

XVIII. Demonstration of KPP Activity

Generally, protein kinase activity is measured by quantifying the phosphorylation of a protein substrate by KPP in the presence of [γ-^{32}P]ATP. KPP is incubated with the protein substrate, ^{32}P-ATP, and an appropriate kinase buffer. The ^{32}P incorporated into the substrate is separated from free ^{32}P-ATP by electrophoresis and the incorporated ^{32}P is counted using a radioisotope counter. The amount of incorporated ^{32}P is proportional to the activity of KPP. A determination of the specific amino acid residue phosphorylated is made by phosphoamino acid analysis of the hydrolyzed protein.

In one alternative, protein kinase activity is measured by quantifying the transfer of gamma phosphate from adenosine triphosphate (ATP) to a serine, threonine or tyrosine residue in a protein substrate. The reaction occurs between a protein kinase sample with a biotinylated peptide substrate and gamma ^{32}P-ATP. Following the reaction, free avidin in solution is added for binding to the biotinylated ^{32}P-peptide product. The binding sample then undergoes a centrifugal ultrafiltration process with a membrane which will retain the product-avidin complex and allow passage of free gamma ^{32}P-ATP. The reservoir of the centrifuged unit containing the ^{32}P-peptide product as retentate is then counted in a scintillation counter. This procedure allows the assay of any type of protein kinase sample, depending on the peptide substrate and kinase reaction buffer selected. This assay is provided in kit form (ASUA, Affinity Ultrafiltration Separation Assay, Transbio Corporation, Baltimore MD, U.S. Patent No. 5,869,275). Suggested substrates and their respective enzymes include but are not limited to: Histone H1 (Sigma) and p34^{cdc2}kinase, Annexin I, Angiotensin (Sigma) and EGF receptor kinase, Annexin II and *src* kinase, ERK1 & ERK2 substrates and MEK, and myelin basic protein and ERK (Pearson, J.D. et al. (1991) Methods Enzymol. 200:62-81).

In another alternative, protein kinase activity of KPP is demonstrated in an assay containing KPP, 50 μl of kinase buffer, 1 μg substrate, such as myelin basic protein (MBP) or synthetic peptide substrates, 1 mM DTT, 10 μg ATP, and 0.5 μCi [γ-^{32}P]ATP. The reaction is incubated at 30°C for 30 minutes and stopped by pipetting onto P81 paper. The unincorporated [γ-^{32}P]ATP is removed by washing and the incorporated radioactivity is measured using a scintillation counter. Alternatively, the reaction is stopped by heating to 100°C in the presence of SDS loading buffer and resolved on a 12% SDS polyacrylamide gel followed by autoradiography. The amount of incorporated ^{32}P is proportional to the activity of KPP.

In yet another alternative, adenylate kinase or guanylate kinase activity of KPP may be measured by the incorporation of ^{32}P from [γ-^{32}P]ATP into ADP or GDP using a gamma radioisotope counter. KPP, in a kinase buffer, is incubated together with the appropriate nucleotide mono-phosphate substrate (AMP or GMP) and ^{32}P-labeled ATP as the phosphate donor. The

reaction is incubated at 37°C and terminated by addition of trichloroacetic acid. The acid extract is neutralized and subjected to gel electrophoresis to separate the mono-, di-, and triphosphonucleotide fractions. The diphosphonucleotide fraction is excised and counted. The radioactivity recovered is proportional to the activity of KPP.

In yet another alternative, other assays for KPP include scintillation proximity assays (SPA), scintillation plate technology and filter binding assays. Useful substrates include recombinant proteins tagged with glutathione transferase, or synthetic peptide substrates tagged with biotin. Inhibitors of KPP activity, such as small organic molecules, proteins or peptides, may be identified by such assays.

In another alternative, phosphatase activity of KPP is measured by the hydrolysis of para-nitrophenyl phosphate (PNPP). KPP is incubated together with PNPP in HEPES buffer pH 7.5, in the presence of 0.1% β-mercaptoethanol at 37°C for 60 min. The reaction is stopped by the addition of 6 ml of 10 N NaOH (Diamond, R.H. et al. (1994) Mol. Cell. Biol. 14:3752-62). Alternatively, acid phosphatase activity of KPP is demonstrated by incubating KPP-containing extract with 100 μl of 10 mM PNPP in 0.1 M sodium citrate, pH 4.5, and 50 μl of 40 mM NaCl at 37°C for 20 min. The reaction is stopped by the addition of 0.5 ml of 0.4 M glycine/NaOH, pH 10.4 (Saftig, P. et al. (1997) J. Biol. Chem. 272:18628-18635). The increase in light absorbance at 410 nm resulting from the hydrolysis of PNPP is measured using a spectrophotometer. The increase in light absorbance is proportional to the activity of KPP in the assay.

In the alternative, KPP activity is determined by measuring the amount of phosphate removed from a phosphorylated protein substrate. Reactions are performed with 2 or 4 nM KPP in a final volume of 30 μl containing 60 mM Tris, pH 7.6, 1 mM EDTA, 1 mM EGTA, 0.1% β-mercaptoethanol and 10 μM substrate, ^{32}P-labeled on serine/threonine or tyrosine, as appropriate. Reactions are initiated with substrate and incubated at 30° C for 10-15 min. Reactions are quenched with 450 μl of 4% (w/v) activated charcoal in 0.6 M HCl, 90 mM $Na_4P_2O_7$, and 2 mM NaH_2PO_4, then centrifuged at 12,000 × *g* for 5 min. Acid-soluble ^{32}Pi is quantified by liquid scintillation counting (Sinclair, C. et al. (1999) J. Biol. Chem. 274:23666-23672).

XIX. Kinase Binding Assay

Binding of KPP to a FLAG-CD44 cyt fusion protein can be determined by incubating KPP with anti-KPP-conjugated immunoaffinity beads followed by incubating portions of the beads (having 10-20 ng of protein) with 0.5 ml of a binding buffer (20 mM Tris-HCL (pH 7.4), 150 mM NaCl, 0.1% bovine serum albumin, and 0.05% Triton X-100) in the presence of ^{125}I-labeled FLAG-CD44cyt fusion protein (5,000 cpm/ng protein) at 4 °C for 5 hours. Following binding, beads were washed thoroughly in the binding buffer and the bead-bound radioactivity measured in a scintillation counter (Bourguignon, L.Y.W. et al. (2001) J. Biol. Chem. 276:7327-7336). The amount of incorporated ^{32}P is proportional to the amount of bound KPP.

XX. Identification of KPP Inhibitors

Compounds to be tested are arrayed in the wells of a 384-well plate in varying concentrations along with an appropriate buffer and substrate, as described in the assays in Example XVII. KPP activity is measured for each well and the ability of each compound to inhibit KPP activity can be determined, as well as the dose-response kinetics. This assay could also be used to identify molecules which enhance KPP activity.

XXI. Identification of KPP Substrates

A KPP "substrate-trapping" assay takes advantage of the increased substrate affinity that may be conferred by certain mutations in the PTP signature sequence of protein tyrosine phosphatases. KPP bearing these mutations form a stable complex with their substrate; this complex may be isolated biochemically. Site-directed mutagenesis of invariant residues in the PTP signature sequence in a clone encoding the catalytic domain of KPP is performed using a method standard in the art or a commercial kit, such as the MUTA-GENE kit from BIO-RAD. For expression of KPP mutants in *Escherichia coli*, DNA fragments containing the mutation are exchanged with the corresponding wild-type sequence in an expression vector bearing the sequence encoding KPP or a glutathione S-transferase (GST)-KPP fusion protein. KPP mutants are expressed in *E. coli* and purified by chromatography.

The expression vector is transfected into COS1 or 293 cells via calcium phosphate-mediated transfection with 20 μg of CsCl-purified DNA per 10-cm dish of cells or 8 μg per 6-cm dish. Forty-eight hours after transfection, cells are stimulated with 100 ng/ml epidermal growth factor to increase tyrosine phosphorylation in cells, as the tyrosine kinase EGFR is abundant in COS cells. Cells are lysed in 50 mM Tris·HCl, pH 7.5/5 mM EDTA/150 mM NaCl/1% Triton X-100/5 mM iodoacetic acid/10 mM sodium phosphate/10 mM NaF/5 μg/ml leupeptin/5 μg/ml aprotinin/1 mM benzamidine (1 ml per 10-cm dish, 0.5 ml per 6-cm dish). KPP is immunoprecipitated from lysates with an appropriate antibody. GST-KPP fusion proteins are precipitated with glutathione-Sepharose, 4 μg of mAb or 10 μl of beads respectively per mg of cell lysate. Complexes can be visualized by PAGE or further purified to identify substrate molecules (Flint, A.J. et al. (1997) Proc. Natl. Acad. Sci. USA 94:1680-1685).

Various modifications and variations of the described compositions, methods, and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. It will be appreciated that the invention provides novel and useful proteins, and their encoding polynucleotides, which can be used in the drug discovery process, as well as methods for using these compositions for the detection, diagnosis, and treatment of diseases and conditions. Although the invention has been described in connection with certain embodiments, it should be

understood that the invention as claimed should not be unduly limited to such specific embodiments. Nor should the description of such embodiments be considered exhaustive or limit the invention to the precise forms disclosed. Furthermore, elements from one embodiment can be readily recombined with elements from one or more other embodiments. Such combinations can form a number of embodiments within the scope of the invention. It is intended that the scope of the invention be defined by the following claims and their equivalents.

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:2-4, SEQ ID NO:8-13 and SEQ ID NO:15,

c) a polypeptide comprising a naturally occurring amino acid sequence at least 97% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:5 and SEQ ID NO:6,

d) a polypeptide comprising a naturally occurring amino acid sequence at least 94% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:7 and SEQ ID NO:1,

e) a polypeptide consisting essentially of a naturally occurring amino acid sequence at least 90% identical to the amino acid sequence of SEQ ID NO:14,

f) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and

g) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

e) an RNA equivalent of a)-d).

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-19 and SEQ ID NO:21-25,

c) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 99% identical to the polynucleotide sequence of SEQ ID NO:29,

d) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 96% identical to the polynucleotide sequence of SEQ ID NO:28,

e) a polynucleotide comprising a naturally occurring polynucleotide sequence at least

93% identical to the polynucleotide sequence of SEQ ID NO:20,

f) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 92% identical to the polynucleotide sequence of SEQ ID NO:27,

g) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 91% identical to the polynucleotide sequence of SEQ ID NO:26,

h) a polynucleotide consisting essentially of a naturally occurring polynucleotide sequence at least 90% identical to the polynucleotide sequence of SEQ ID NO:30,

i) a polynucleotide complementary to a polynucleotide of a),

j) a polynucleotide complementary to a polynucleotide of b),

k) a polynucleotide complementary to a polynucleotide of c),

l) a polynucleotide complementary to a polynucleotide of d),

m) a polynucleotide complementary to a polynucleotide of e),

n) a polynucleotide complementary to a polynucleotide of f),

o) a polynucleotide complementary to a polynucleotide of g),

p) a polynucleotide complementary to a polynucleotide of h), and

q) an RNA equivalent of a)-p).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain

reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.

19. A method for treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) contacting a sample comprising a polypeptide of claim 1 with a compound, and

b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) contacting a sample comprising a polypeptide of claim 1 with a compound, and

b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional

KPP, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) contacting a sample comprising the target polynucleotide with a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of screening for potential toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target

polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample indicates potential toxicity of the test compound.

30. A method for a diagnostic test for a condition or disease associated with the expression of KPP in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of KPP in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, further comprising a label.

35. A method of diagnosing a condition or disease associated with the expression of KPP in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim

11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibodies from the animal, and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15 in a sample, the method comprising:

a) incubating the antibody of claim 11 with the sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15 from a sample, the method comprising:

a) incubating the antibody of claim 11 with the sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

71. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:16.

72. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:17.

73. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:18.

74. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:19.

75. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:20.

76. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:21.

77. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:22.

78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID

NO:23.

79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.

80. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:25.

81. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:26.

82. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:27.

83. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:28.

84. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:29.

85. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:30.

ABSTRACT OF THE DISCLOSURE

Various embodiments of the invention provide human kinases and phosphatases (KPP) and polynucleotides which identify and encode KPP. Embodiments of the invention also provide expression vectors, host cells, antibodies, agonists, and antagonists. Other embodiments provide methods for diagnosing, treating, or preventing disorders associated with aberrant expression of KPP.

<110> CHAWLA, Narinder K.
BECHA, Shanya D.
WILSON, Amy D.
JIN, Pei

<120> KINASES AND PHOSPHATASES

<130> PF-1724 P

<140> To Be Assigned
<141> Herewith

<160> 30
<170> PERL Program

<210> 1
<211> 157
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526185CD1

<400> 1
Met Ala His Ser Pro Val Gln Ser Gly Leu Pro Gly Met Gln Asn
1 5 10 15
Leu Lys Ala Asp Pro Glu Glu Leu Phe Thr Lys Leu Glu Lys Ile
20 25 30
Gly Lys Gly Ser Phe Gly Glu Val Phe Lys Gly Ile Asp Asn Arg
35 40 45
Thr Gln Lys Val Val Ala Ile Lys Ile Ile Asp Leu Glu Glu Ala
50 55 60
Glu Asp Glu Ile Glu Asp Ile Gln Gln Glu Ile Thr Val Leu Ser
65 70 75
Gln Cys Asp Ser Pro Tyr Val Thr Lys Tyr Tyr Gly Ser Tyr Leu
80 85 90
Lys Asp Thr Lys Leu Trp Ile Ile Met Glu Tyr Leu Gly Gly Gly
95 100 105
Ser Ala Leu Asp Leu Leu Glu Pro Gly Pro Leu Asp Glu Thr Gln
110 115 120
Ile Ala Thr Ile Leu Arg Glu Ile Leu Lys Gly Leu Asp Tyr Leu
125 130 135
His Ser Glu Lys Lys Ile His Arg Asp Ile Lys Gly Arg His Leu
140 145 150
Val Pro Gly His Asn Ser Tyr
155

<210> 2
<211> 305
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526192CD1

<400> 2
Met Asp Phe Asp Lys Lys Gly Gly Lys Gly Glu Thr Glu Glu Gly

```
1               5                   10                  15
Arg Arg Met Ser Lys Ala Gly Gly Gly Arg Ser Ser His Gly Ile
                20                  25                  30
Arg Ser Ser Gly Thr Ser Ser Gly Val Leu Met Val Gly Pro Asn
                35                  40                  45
Phe Arg Val Gly Lys Lys Ile Gly Cys Gly Asn Phe Gly Glu Leu
                50                  55                  60
Arg Leu Gly Lys Asn Leu Tyr Thr Asn Glu Tyr Val Ala Ile Lys
                65                  70                  75
Leu Val Ser Arg Pro Leu His Pro Thr Pro Ala Asp Val Pro Pro
                80                  85                  90
Arg Asp Phe Arg Ala Ala Thr Arg Ser Pro Gly Asp Ser Leu Leu
                95                  100                 105
Cys Pro Gln Glu Pro Ile Lys Ser Arg Ala Pro Gln Leu His Leu
                110                 115                 120
Glu Tyr Arg Phe Tyr Lys Gln Leu Ser Ala Thr Glu Gly Val Pro
                125                 130                 135
Gln Val Tyr Tyr Phe Gly Pro Cys Gly Lys Tyr Asn Ala Met Val
                140                 145                 150
Leu Glu Leu Leu Gly Pro Ile Leu Glu Asp Leu Phe Asp Leu Cys
                155                 160                 165
Asp Arg Thr Phe Thr Leu Thr Thr Val Leu Met Ile Ala Ile Gln
                170                 175                 180
Leu Ile Thr Arg Met Glu Tyr Val His Thr Lys Ser Leu Ile Tyr
                185                 190                 195
Arg Asp Val Lys Pro Glu Asn Phe Leu Val Gly Arg Pro Gly Thr
                200                 205                 210
Lys Arg Gln His Ala Ile His Ile Ile Asp Phe Gly Leu Ala Lys
                215                 220                 225
Glu Tyr Ile Asp Pro Glu Thr Lys Lys His Ile Pro Tyr Arg Glu
                230                 235                 240
His Lys Ser Leu Thr Gly Thr Ala Arg Tyr Met Ser Ile Asn Thr
                245                 250                 255
His Leu Gly Lys Glu Gln Ser Arg Arg Asp Asp Leu Glu Ala Leu
                260                 265                 270
Gly His Met Phe Met Tyr Phe Leu Arg Gly Ser Leu Pro Trp Gln
                275                 280                 285
Gly Leu Lys Val Gly Glu Glu Ala Gly Gln Ala Gly Gly Asp Ala
                290                 295                 300
Gly Arg Glu Gln Gly
                305
```

<210> 3
<211> 930
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526193CD1

<400> 3

```
Met Lys Lys Phe Phe Asp Ser Arg Arg Glu Gln Gly Gly Ser Gly
1               5                   10                  15
Leu Gly Ser Gly Ser Ser Gly Gly Gly Gly Ser Thr Ser Gly Leu
                20                  25                  30
Gly Ser Gly Tyr Ile Gly Arg Val Phe Gly Ile Gly Arg Gln Gln
                35                  40                  45
Val Thr Val Asp Glu Val Leu Ala Glu Gly Gly Phe Ala Ile Val
```

PF-1724 P

50 55 60
Phe Leu Val Arg Thr Ser Asn Gly Met Lys Cys Ala Leu Lys Arg
65 70 75
Met Phe Val Asn Asn Glu His Asp Leu Gln Val Cys Lys Arg Glu
80 85 90
Ile Gln Ile Met Arg Asp Leu Ser Gly His Lys Asn Ile Val Gly
95 100 105
Tyr Ile Asp Ser Ser Ile Asn Asn Val Ser Ser Gly Asp Val Trp
110 115 120
Glu Val Leu Ile Leu Met Asp Phe Cys Arg Gly Gly Gln Val Val
125 130 135
Asn Leu Met Asn Gln Arg Leu Gln Thr Gly Phe Thr Glu Asn Glu
140 145 150
Val Leu Gln Ile Phe Cys Asp Thr Cys Glu Ala Val Ala Arg Leu
155 160 165
His Gln Cys Lys Thr Pro Ile Ile His Arg Asp Leu Lys Val Glu
170 175 180
Asn Ile Leu Leu His Asp Arg Gly His Tyr Val Leu Cys Asp Phe
185 190 195
Gly Ser Ala Thr Asn Lys Phe Gln Asn Pro Gln Thr Glu Gly Val
200 205 210
Asn Ala Val Glu Asp Glu Ile Lys Lys Tyr Thr Thr Leu Ser Tyr
215 220 225
Arg Ala Pro Glu Met Val Asn Leu Tyr Ser Gly Lys Ile Ile Thr
230 235 240
Thr Lys Ala Asp Ile Trp Ala Leu Gly Cys Leu Leu Tyr Lys Leu
245 250 255
Cys Tyr Phe Thr Leu Pro Phe Gly Glu Ser Gln Val Ala Ile Cys
260 265 270
Asp Gly Asn Phe Thr Ile Pro Asp Asn Ser Arg Tyr Ser Gln Asp
275 280 285
Met His Cys Leu Ile Arg Tyr Met Leu Glu Pro Asp Pro Asp Lys
290 295 300
Arg Pro Asp Ile Tyr Gln Val Ser Tyr Phe Ser Phe Lys Leu Leu
305 310 315
Lys Lys Glu Cys Pro Ile Pro Asn Val Gln Asn Ser Pro Ile Pro
320 325 330
Ala Lys Leu Pro Glu Pro Val Lys Ala Ser Glu Ala Ala Ala Lys
335 340 345
Lys Thr Gln Pro Lys Ala Arg Leu Thr Asp Pro Ile Pro Thr Thr
350 355 360
Glu Thr Ser Ile Ala Pro Arg Gln Arg Pro Lys Ala Gly Gln Thr
365 370 375
Gln Pro Asn Pro Gly Ile Leu Pro Ile Gln Pro Ala Leu Thr Pro
380 385 390
Arg Lys Arg Ala Thr Val Gln Pro Pro Pro Gln Ala Ala Gly Ser
395 400 405
Ser Asn Gln Pro Gly Leu Leu Ala Ser Val Pro Gln Pro Lys Pro
410 415 420
Gln Ala Pro Pro Ser Gln Pro Leu Pro Gln Thr Gln Ala Lys Gln
425 430 435
Pro Gln Ala Pro Pro Thr Pro Gln Gln Thr Pro Ser Thr Gln Ala
440 445 450
Gln Gly Leu Pro Ala Gln Ala Gln Ala Thr Pro Gln His Gln Gln
455 460 465
Gln Leu Phe Leu Lys Gln Gln Gln Gln Gln Gln Gln Pro Pro Pro
470 475 480
Ala Gln Gln Gln Pro Ala Gly Thr Phe Tyr Gln Gln Gln Gln Ala
485 490 495

Gln Thr Gln Gln Phe Gln Ala Val His Pro Ala Thr Gln Gln Pro
500 505 510
Ala Ile Ala Gln Phe Pro Val Val Ser Gln Gly Gly Ser Gln Gln
515 520 525
Gln Leu Met Gln Asn Phe Tyr Gln Gln Gln Gln Gln Gln Gln Gln
530 535 540
Gln Gln Gln Gln Gln Gln Leu Ala Thr Ala Leu His Gln Gln Gln
545 550 555
Leu Met Thr Gln Gln Ala Ala Leu Gln Gln Lys Pro Thr Met Ala
560 565 570
Ala Gly Gln Gln Pro Gln Pro Gln Pro Ala Ala Ala Pro Gln Pro
575 580 585
Ala Pro Ala Gln Glu Pro Ala Gln Ile Gln Ala Pro Val Arg Gln
590 595 600
Gln Pro Lys Val Gln Thr Thr Pro Pro Pro Ala Val Gln Gly Gln
605 610 615
Lys Val Gly Ser Leu Thr Pro Pro Ser Ser Pro Lys Thr Gln Arg
620 625 630
Ala Gly His Arg Arg Ile Leu Ser Asp Val Thr His Ser Ala Val
635 640 645
Phe Gly Val Pro Ala Ser Lys Ser Thr Gln Leu Leu Gln Ala Ala
650 655 660
Ala Ala Glu Ala Ser Leu Asn Lys Ser Lys Ser Ala Thr Thr Thr
665 670 675
Pro Ser Gly Ser Pro Arg Thr Ser Gln Gln Asn Val Tyr Asn Pro
680 685 690
Ser Glu Gly Ser Thr Trp Asn Pro Phe Asp Asp Asp Asn Phe Ser
695 700 705
Lys Leu Thr Ala Glu Glu Leu Leu Asn Lys Asp Phe Ala Lys Leu
710 715 720
Gly Glu Gly Lys His Pro Glu Lys Leu Gly Gly Ser Ala Glu Ser
725 730 735
Leu Ile Pro Gly Phe Gln Ser Thr Gln Gly Asp Ala Phe Ala Thr
740 745 750
Thr Ser Phe Ser Ala Gly Thr Glu Lys Leu Ile Glu Gly Leu Lys
755 760 765
Ser Pro Asp Thr Ser Leu Leu Leu Pro Asp Leu Leu Pro Met Thr
770 775 780
Asp Pro Phe Gly Ser Thr Ser Asp Ala Val Ile Glu Lys Ala Asp
785 790 795
Val Ala Val Glu Ser Leu Ile Pro Gly Leu Glu Pro Pro Val Pro
800 805 810
Gln Arg Leu Pro Ser Gln Thr Glu Ser Val Thr Ser Asn Arg Thr
815 820 825
Asp Ser Leu Thr Gly Glu Asp Ser Leu Leu Asp Cys Ser Leu Leu
830 835 840
Ser Asn Pro Thr Thr Asp Leu Leu Glu Glu Phe Ala Pro Thr Ala
845 850 855
Ile Ser Ala Pro Val His Lys Ala Ala Glu Asp Ser Asn Leu Ile
860 865 870
Ser Gly Phe Asp Val Pro Glu Gly Ser Asp Lys Val Ala Glu Asp
875 880 885
Glu Phe Asp Pro Ile Pro Val Leu Ile Thr Lys Asn Pro Gln Gly
890 895 900
Gly His Ser Arg Asn Ser Ser Gly Ser Ser Glu Ser Ser Leu Pro
905 910 915
Asn Leu Ala Arg Ser Leu Leu Leu Val Asp Gln Leu Ile Asp Leu
920 925 930

<210> 4
<211> 118
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526196CD1

<400> 4
Met Ser Leu Leu Gln Ser Ala Leu Asp Phe Leu Ala Gly Pro Gly
1 5 10 15
Ser Leu Gly Gly Ala Ser Gly Arg Asp Gln Ser Asp Phe Val Gly
20 25 30
Gln Thr Val Glu Leu Gly Glu Leu Arg Leu Arg Val Arg Arg Val
35 40 45
Leu Ala Glu Gly Gly Phe Ala Phe Val Tyr Glu Ala Gln Asp Val
50 55 60
Gly Ser Gly Arg Glu Tyr Ala Leu Lys Arg Leu Leu Ser Asn Glu
65 70 75
Glu Glu Lys Asn Arg Ala Ile Ile Gln Glu Val Cys Phe Met Leu
80 85 90
Cys Ser Leu Gly Glu Pro Ala Gly Cys Leu Ser Val Gly Ser Gly
95 100 105
Gly His Ser His Ala Ser Ala Ser Leu Arg Thr Ala Pro
110 115

<210> 5
<211> 1355
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526198CD1

<400> 5
Met Ser Leu Leu Gln Ser Ala Leu Asp Phe Leu Ala Gly Pro Gly
1 5 10 15
Ser Leu Gly Gly Ala Ser Gly Arg Asp Gln Ser Asp Phe Val Gly
20 25 30
Gln Thr Val Glu Leu Gly Glu Leu Arg Leu Arg Val Arg Arg Val
35 40 45
Leu Ala Glu Gly Gly Phe Ala Phe Val Tyr Glu Ala Gln Asp Val
50 55 60
Gly Ser Gly Arg Glu Tyr Ala Leu Lys Arg Leu Leu Ser Asn Glu
65 70 75
Glu Glu Lys Asn Arg Ala Ile Ile Gln Glu Val Cys Phe Met Lys
80 85 90
Lys Leu Ser Gly His Pro Asn Ile Val Gln Phe Cys Ser Ala Ala
95 100 105
Ser Ile Gly Lys Glu Glu Ser Asp Thr Gly Gln Ala Glu Phe Leu
110 115 120
Leu Leu Thr Glu Leu Cys Lys Gly Gln Leu Val Glu Phe Leu Lys
125 130 135
Lys Met Glu Ser Arg Gly Pro Leu Ser Cys Asp Thr Val Leu Lys
140 145 150
Ile Phe Tyr Gln Thr Cys Arg Ala Val Gln His Met His Arg Gln
155 160 165

Lys Pro Pro Ile Ile His Arg Asp Leu Lys Val Glu Asn Leu Leu
170 175 180
Leu Ser Asn Gln Gly Thr Ile Lys Leu Cys Asp Phe Gly Ser Ala
185 190 195
Thr Thr Ile Ser His Tyr Pro Asp Tyr Ser Trp Ser Ala Gln Arg
200 205 210
Arg Ala Leu Val Glu Glu Glu Ile Thr Arg Asn Thr Thr Pro Met
215 220 225
Tyr Arg Thr Pro Glu Ile Ile Asp Leu Tyr Ser Asn Phe Pro Ile
230 235 240
Gly Glu Lys Gln Asp Ile Trp Ala Leu Gly Cys Ile Leu Tyr Leu
245 250 255
Leu Cys Phe Arg Gln His Pro Phe Glu Asp Gly Ala Lys Leu Arg
260 265 270
Ile Val Asn Gly Lys Tyr Ser Ile Pro Pro His Asp Thr Gln Tyr
275 280 285
Thr Val Phe His Ser Leu Ile Arg Ala Met Leu Gln Val Asn Pro
290 295 300
Glu Glu Arg Leu Ser Ile Ala Glu Val Val His Gln Leu Gln Glu
305 310 315
Ile Ala Ala Ala Arg Asn Val Asn Pro Lys Ser Pro Ile Thr Glu
320 325 330
Leu Leu Glu Gln Asn Gly Gly Tyr Gly Ser Ala Thr Leu Ser Arg
335 340 345
Gly Pro Pro Pro Pro Val Gly Pro Ala Gly Ser Gly Tyr Ser Gly
350 355 360
Gly Leu Ala Leu Ala Glu Tyr Asp Gln Pro Tyr Gly Gly Phe Leu
365 370 375
Asp Ile Leu Arg Gly Gly Thr Glu Arg Leu Phe Thr Asn Leu Lys
380 385 390
Asp Thr Ser Ser Lys Val Ile Gln Ser Val Ala Asn Tyr Ala Lys
395 400 405
Gly Asp Leu Asp Ile Ser Tyr Ile Thr Ser Arg Ile Ala Val Met
410 415 420
Ser Phe Pro Ala Glu Gly Val Glu Ser Ala Leu Lys Asn Asn Ile
425 430 435
Glu Asp Val Arg Leu Phe Leu Asp Ser Lys His Pro Gly His Tyr
440 445 450
Ala Val Tyr Asn Leu Ser Pro Arg Thr Tyr Arg Pro Ser Arg Phe
455 460 465
His Asn Arg Val Ser Glu Cys Gly Trp Ala Ala Arg Arg Ala Pro
470 475 480
His Leu His Thr Leu Tyr Asn Ile Cys Arg Asn Met His Ala Trp
485 490 495
Leu Arg Gln Asp His Lys Asn Val Cys Val Val His Cys Met Asp
500 505 510
Gly Arg Ala Ala Ser Ala Val Ala Val Cys Ser Phe Leu Cys Phe
515 520 525
Cys Arg Leu Phe Ser Thr Ala Glu Ala Ala Val Tyr Met Phe Ser
530 535 540
Met Lys Arg Cys Pro Pro Gly Ile Trp Pro Ser His Lys Arg Tyr
545 550 555
Ile Glu Tyr Met Cys Asp Met Val Ala Glu Glu Pro Ile Thr Pro
560 565 570
His Ser Lys Pro Ile Leu Val Arg Ala Val Val Met Thr Pro Val
575 580 585
Pro Leu Phe Ser Lys Gln Arg Ser Gly Cys Arg Pro Phe Cys Glu
590 595 600
Val Tyr Val Gly Asp Glu Arg Val Ala Ser Thr Ser Gln Glu Tyr

605 610 615
Asp Lys Met Arg Asp Phe Lys Ile Glu Asp Gly Ile Ala Val Ile
620 625 630
Pro Leu Gly Val Thr Val Gln Gly Asp Val Leu Ile Val Ile Tyr
635 640 645
His Ala Arg Ser Thr Leu Gly Gly Arg Leu Gln Ala Lys Met Ala
650 655 660
Ser Met Lys Met Phe Gln Ile Gln Phe His Thr Gly Phe Val Pro
665 670 675
Arg Asn Ala Thr Thr Val Lys Phe Ala Lys Tyr Asp Leu Asp Ala
680 685 690
Cys Asp Ile Gln Glu Lys Tyr Pro Asp Leu Phe Gln Val Asn Leu
695 700 705
Glu Val Glu Val Glu Pro Arg Asp Arg Pro Ser Arg Glu Ala Pro
710 715 720
Pro Trp Glu Asn Ser Ser Met Arg Gly Leu Asn Pro Lys Ile Leu
725 730 735
Phe Ser Ser Arg Glu Glu Gln Gln Asp Ile Leu Ser Lys Phe Gly
740 745 750
Lys Pro Glu Leu Pro Arg Gln Pro Gly Ser Thr Ala Gln Tyr Asp
755 760 765
Ala Gly Ala Gly Ser Pro Glu Ala Glu Pro Thr Asp Ser Asp Ser
770 775 780
Pro Pro Ser Ser Ser Ala Asp Ala Ser Arg Phe Leu His Thr Leu
785 790 795
Asp Trp Gln Glu Glu Lys Glu Ala Glu Thr Gly Ala Glu Asn Ala
800 805 810
Ser Ser Lys Glu Ser Glu Ser Ala Leu Met Glu Asp Arg Asp Glu
815 820 825
Ser Glu Val Ser Asp Glu Gly Gly Ser Pro Ile Ser Ser Glu Gly
830 835 840
Gln Glu Pro Arg Ala Asp Pro Glu Pro Pro Gly Leu Ala Ala Gly
845 850 855
Leu Val Gln Gln Asp Leu Val Phe Glu Val Glu Thr Pro Ala Val
860 865 870
Leu Pro Glu Pro Val Pro Gln Glu Asp Gly Val Asp Leu Leu Gly
875 880 885
Leu His Ser Glu Val Gly Ala Gly Pro Ala Val Pro Pro Gln Ala
890 895 900
Cys Lys Ala Pro Ser Ser Asn Thr Asp Leu Leu Ser Cys Leu Leu
905 910 915
Gly Pro Pro Glu Ala Ala Ser Gln Gly Pro Pro Glu Asp Leu Leu
920 925 930
Ser Glu Asp Pro Leu Leu Leu Ala Ser Pro Ala Pro Pro Leu Ser
935 940 945
Val Gln Ser Thr Pro Arg Gly Gly Pro Pro Ala Ala Ala Asp Pro
950 955 960
Phe Gly Pro Leu Leu Pro Ser Ser Gly Asn Asn Ser Gln Pro Cys
965 970 975
Ser Asn Pro Asp Leu Phe Gly Glu Phe Leu Asn Ser Asp Ser Val
980 985 990
Thr Val Pro Pro Ser Phe Pro Ser Ala His Ser Ala Pro Pro Pro
995 1000 1005
Ser Cys Ser Ala Asp Phe Leu His Leu Gly Asp Leu Pro Gly Glu
1010 1015 1020
Pro Ser Lys Met Thr Ala Ser Ser Ser Asn Pro Asp Leu Leu Gly
1025 1030 1035
Gly Trp Ala Ala Trp Thr Glu Thr Ala Ala Ser Ala Val Ala Pro
1040 1045 1050

Thr Pro Ala Thr Glu Gly Pro Leu Phe Ser Pro Gly Gly Gln Pro
1055 1060 1065
Ala Pro Cys Gly Ser Gln Ala Ser Trp Thr Lys Ser Gln Asn Pro
1070 1075 1080
Asp Pro Phe Ala Asp Leu Gly Asp Leu Ser Ser Gly Leu Gln Asp
1085 1090 1095
Pro Gln Ala Gln Ser Thr Val Ser Pro Arg Gly Gln Arg Val Cys
1100 1105 1110
Thr Cys Ser Arg Arg Leu Pro Thr Gly Lys Leu Lys Pro Gly Val
1115 1120 1125
Ala Asp Thr Gly Thr Ala Ala Ser Pro His Arg His Cys Gly Ser
1130 1135 1140
Pro Ala Gly Phe Pro Pro Gly Gly Phe Ile Pro Lys Thr Ala Thr
1145 1150 1155
Thr Pro Lys Gly Ser Ser Ser Trp Gln Thr Ser Arg Pro Pro Ala
1160 1165 1170
Gln Gly Ala Ser Trp Pro Pro Gln Ala Lys Pro Pro Pro Lys Ala
1175 1180 1185
Cys Thr Gln Pro Arg Pro Asn Tyr Ala Ser Asn Phe Ser Val Ile
1190 1195 1200
Gly Ala Arg Glu Glu Arg Gly Val Arg Ala Pro Ser Phe Ala Gln
1205 1210 1215
Lys Pro Lys Val Ser Glu Asn Asp Phe Glu Asp Leu Leu Ser Asn
1220 1225 1230
Gln Gly Phe Ser Ser Arg Ser Asp Lys Lys Gly Pro Lys Thr Ile
1235 1240 1245
Ala Glu Met Arg Lys Gln Asp Leu Ala Lys Asp Thr Asp Pro Leu
1250 1255 1260
Lys Leu Lys Leu Leu Asp Trp Ile Glu Gly Lys Glu Arg Asn Ile
1265 1270 1275
Arg Ala Leu Leu Ser Thr Leu His Thr Val Leu Trp Asp Gly Glu
1280 1285 1290
Ser Arg Trp Thr Pro Val Gly Met Ala Asp Leu Val Ala Pro Glu
1295 1300 1305
Gln Val Lys Lys His Tyr Arg Arg Ala Val Leu Ala Val His Pro
1310 1315 1320
Asp Lys Ala Ala Gly Gln Pro Tyr Glu Gln His Ala Lys Met Ile
1325 1330 1335
Phe Met Glu Leu Asn Asp Ala Trp Ser Glu Phe Glu Asn Gln Gly
1340 1345 1350
Ser Arg Pro Leu Phe
1355

<210> 6
<211> 490
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526208CD1

<400> 6
Met Ala Ser Thr Thr Thr Cys Thr Arg Phe Thr Asp Glu Tyr Gln
1 5 10 15
Leu Phe Glu Glu Leu Gly Lys Gly Ala Phe Ser Val Val Arg Arg
20 25 30
Cys Met Lys Ile Pro Thr Gly Gln Glu Tyr Ala Ala Lys Ile Ile
35 40 45

```
Asn Thr Lys Lys Leu Ser Ala Arg Val Arg Leu His Asp Ser Ile
                50                  55                  60
Ser Glu Glu Gly Phe His Tyr Leu Val Phe Asp Leu Val Thr Gly
                65                  70                  75
Gly Glu Leu Phe Glu Asp Ile Val Ala Arg Glu Tyr Tyr Ser Glu
                80                  85                  90
Ala Asp Ala Ser His Cys Ile Gln Gln Ile Leu Glu Ala Val Leu
                95                  100                 105
His Cys His Gln Met Gly Val Val His Arg Asp Leu Lys Pro Glu
                110                 115                 120
Asn Leu Leu Leu Ala Ser Lys Ser Lys Gly Ala Ala Val Lys Leu
                125                 130                 135
Ala Asp Phe Gly Leu Ala Ile Glu Val Gln Gly Asp Gln Gln Ala
                140                 145                 150
Trp Phe Gly Phe Ala Gly Thr Pro Gly Tyr Leu Ser Pro Glu Val
                155                 160                 165
Leu Arg Lys Asp Pro Tyr Gly Lys Pro Val Asp Met Trp Ala Cys
                170                 175                 180
Gly Val Ile Leu Tyr Ile Leu Leu Val Gly Tyr Pro Pro Phe Trp
                185                 190                 195
Asp Glu Asp Gln His Arg Leu Tyr Gln Gln Ile Lys Ala Gly Ala
                200                 205                 210
Tyr Asp Phe Pro Ser Pro Glu Trp Asp Thr Val Thr Pro Glu Ala
                215                 220                 225
Lys Asp Leu Ile Asn Lys Met Leu Thr Ile Asn Pro Ala Lys Arg
                230                 235                 240
Ile Thr Ala Ser Glu Ala Leu Lys His Pro Trp Ile Cys Gln Arg
                245                 250                 255
Ser Thr Val Ala Ser Met Met His Arg Gln Glu Thr Val Asp Cys
                260                 265                 270
Leu Lys Lys Phe Asn Ala Arg Arg Lys Leu Lys Gly Ala Ile Leu
                275                 280                 285
Thr Thr Met Leu Ala Thr Arg Asn Phe Ser Ala Ala Lys Ser Leu
                290                 295                 300
Leu Lys Lys Pro Asp Gly Val Lys Lys Arg Lys Ser Ser Ser Ser
                305                 310                 315
Val Gln Met Met Glu Ser Thr Glu Ser Ser Asn Thr Thr Ile Glu
                320                 325                 330
Asp Glu Asp Val Glu Ala Arg Lys Gln Glu Ile Ile Lys Val Thr
                335                 340                 345
Glu Gln Leu Ile Glu Ala Ile Asn Asn Gly Asp Phe Glu Ala Tyr
                350                 355                 360
Thr Lys Ile Cys Asp Pro Gly Leu Thr Ala Phe Glu Pro Glu Ala
                365                 370                 375
Leu Gly Asn Leu Val Glu Gly Met Asp Phe His Arg Phe Tyr Phe
                380                 385                 390
Glu Asn Ala Leu Ser Lys Ser Asn Lys Pro Ile His Thr Ile Ile
                395                 400                 405
Leu Asn Pro His Val His Leu Val Gly Asp Asp Ala Ala Cys Ile
                410                 415                 420
Ala Tyr Ile Arg Leu Thr Gln Tyr Met Asp Gly Ser Gly Met Pro
                425                 430                 435
Lys Thr Met Gln Ser Glu Glu Thr Arg Val Trp His Arg Arg Asp
                440                 445                 450
Gly Lys Trp Gln Asn Val His Phe His Arg Ser Gly Ser Pro Thr
                455                 460                 465
Val Pro Ile Lys Pro Pro Cys Ile Pro Asn Gly Lys Glu Asn Phe
                470                 475                 480
Ser Gly Gly Thr Ser Leu Trp Gln Asn Ile
```

485 490

<210> 7
<211> 344
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526212CD1

<400> 7
Met Ala Ser Thr Thr Thr Cys Thr Arg Phe Thr Asp Glu Tyr Gln
1 5 10 15
Leu Phe Glu Glu Leu Gly Lys Gly Ala Phe Ser Val Val Arg Arg
20 25 30
Cys Met Lys Ile Pro Thr Gly Gln Glu Tyr Ala Ala Lys Ile Ile
35 40 45
Asn Thr Lys Lys Leu Ser Ala Arg Val Arg Leu His Asp Ser Ile
50 55 60
Ser Glu Glu Gly Phe His Tyr Leu Val Val Asp Leu Val Thr Gly
65 70 75
Gly Glu Leu Phe Glu Asp Ile Val Ala Arg Glu Tyr Tyr Ser Glu
80 85 90
Ala Asp Ala Ser His Cys Ile Gln Gln Ile Leu Glu Ala Val Leu
95 100 105
His Cys His Gln Met Gly Val Val His Arg Asp Leu Lys Pro Glu
110 115 120
Asn Leu Leu Leu Ala Ser Lys Ser Lys Gly Ala Ala Val Lys Leu
125 130 135
Ala Asp Phe Gly Leu Ala Ile Glu Val Gln Gly Asp Gln Gln Ala
140 145 150
Trp Phe Gly Phe Ala Gly Thr Pro Gly Tyr Leu Ser Pro Glu Val
155 160 165
Leu Arg Lys Asp Pro Tyr Gly Lys Pro Val Asp Met Trp Ala Cys
170 175 180
Gly Val Ile Leu Tyr Ile Leu Leu Val Gly Tyr Pro Pro Phe Trp
185 190 195
Asp Glu Asp Gln His Arg Leu Tyr Gln Gln Ile Lys Ala Gly Ala
200 205 210
Tyr Asp Phe Pro Ser Pro Glu Trp Asp Thr Val Thr Pro Glu Ala
215 220 225
Lys Asp Leu Ile Asn Lys Met Leu Thr Ile Asn Pro Ala Lys Arg
230 235 240
Ile Thr Ala Ser Glu Ala Leu Lys His Pro Trp Ile Cys Gln Arg
245 250 255
Ser Thr Val Ala Ser Met Met His Arg Gln Glu Thr Val Asp Cys
260 265 270
Leu Lys Lys Phe Asn Ala Arg Arg Lys Leu Lys Gly Ala Ile Leu
275 280 285
Thr Thr Met Leu Ala Thr Arg Asn Phe Ser Ala Ala Lys Ser Leu
290 295 300
Leu Lys Lys Pro Asp Gly Val Lys Glu Ser Thr Glu Ser Ser Asn
305 310 315
Thr Thr Ile Glu Asp Glu Asp Val Lys Gly Thr Val Ala His Ala
320 325 330
Cys Asn Pro Ser Thr Leu Gly Gly Arg Gly Gly Gln Ile Thr
335 340

<210> 8
<211> 89
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526213CD1

<400> 8
Met Lys Lys Phe Ser Arg Met Pro Lys Ser Glu Gly Gly Ser Gly
1 5 10 15
Gly Gly Ala Ala Gly Gly Gly Ala Gly Gly Ala Gly Ala Gly Ala
20 25 30
Gly Cys Gly Ser Gly Gly Ser Ser Val Gly Val Arg Val Phe Ala
35 40 45
Val Gly Arg His Gln Val Thr Leu Glu Glu Ser Leu Ala Glu Val
50 55 60
Ile Gln Met Leu Pro Val Gln Glu Pro Arg Leu Glu Tyr Arg Val
65 70 75
Pro Leu Ile Ser Ser Gly Arg Arg Arg Leu Arg Arg Arg Cys
80 85

<210> 9
<211> 88
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526214CD1

<400> 9
Met Lys Lys Phe Ser Arg Met Pro Lys Ser Glu Gly Gly Ser Gly
1 5 10 15
Gly Gly Ala Ala Gly Gly Gly Ala Gly Gly Ala Gly Ala Gly Ala
20 25 30
Gly Cys Gly Ser Gly Gly Ser Ser Val Gly Val Arg Val Phe Ala
35 40 45
Val Gly Arg His Gln Val Thr Leu Glu Glu Ser Leu Ala Glu Gly
50 55 60
Thr Gly Ala Arg Gly Gly Ser Asp Arg Gln Val Asp Ser Pro Gln
65 70 75
Phe Ser Ser Cys Val Leu Thr Val Glu Ser Asp Val His
80 85

<210> 10
<211> 137
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526228CD1

<400> 10
Met Ser Thr Ala Ser Ala Ala Ser Ser Ser Ser Ser Ser Ser Ala
1 5 10 15
Gly Glu Met Ile Glu Ala Pro Ser Gln Val Leu Asn Phe Glu Glu

```
                    20                      25                      30
Ile Asp Tyr Lys Glu Ile Glu Val Glu Glu Val Val Gly Arg Gly
                    35                      40                      45
Ala Phe Gly Val Val Cys Lys Ala Lys Trp Arg Ala Lys Asp Val
                    50                      55                      60
Ala Ile Lys Gln Ile Glu Ser Glu Ser Glu Arg Lys Ala Phe Ile
                    65                      70                      75
Val Glu Leu Arg Gln Leu Ser Arg Val Asn His Pro Asn Ile Val
                    80                      85                      90
Lys Leu Tyr Gly Ala Cys Leu Asn Pro Val Cys Leu Val Met Glu
                    95                      100                     105
Tyr Ala Glu Gly Gly Ser Leu Tyr Asn Val Cys Ala Phe Leu Ser
                    110                     115                     120
Gln Cys Cys Met Val Leu Asn His Cys His Ile Ile Leu Leu Pro
                    125                     130                     135
Thr Gln
```

<210> 11
<211> 243
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526246CD1

<400> 11

```
Met Ala Asp Leu Glu Ala Val Leu Ala Asp Val Ser Tyr Leu Met
  1               5                   10                  15
Ala Met Glu Lys Ser Lys Ala Thr Pro Ala Ala Arg Ala Ser Lys
                 20                   25                  30
Lys Ile Leu Leu Pro Glu Pro Ser Ile Arg Ser Val Met Gln Lys
                 35                   40                  45
Tyr Leu Glu Asp Arg Gly Glu Val Thr Phe Glu Lys Ile Phe Ser
                 50                   55                  60
Gln Lys Leu Gly Tyr Leu Leu Phe Arg Asp Phe Cys Leu Asn His
                 65                   70                  75
Leu Glu Glu Ala Arg Pro Leu Val Glu Phe Tyr Glu Glu Ile Lys
                 80                   85                  90
Lys Tyr Glu Lys Leu Glu Thr Glu Glu Glu Arg Val Ala Arg Ser
                 95                   100                 105
Arg Glu Ile Phe Asp Ser Tyr Ile Met Lys Glu Leu Leu Ala Cys
                 110                  115                 120
Ser His Pro Phe Ser Lys Ser Ala Thr Glu His Val Gln Gly His
                 125                  130                 135
Leu Gly Lys Lys Gln Val Pro Pro Asp Leu Phe Gln Pro Tyr Ile
                 140                  145                 150
Glu Glu Ile Cys Gln Asn Leu Arg Gly Asp Val Phe Gln Lys Phe
                 155                  160                 165
Ile Glu Ser Asp Lys Phe Thr Arg Phe Cys Gln Trp Lys Asn Val
                 170                  175                 180
Glu Leu Asn Ile His Val Ser Gly Leu Gly Trp Gly Met Glu Ser
                 185                  190                 195
His Ala Pro Cys Cys Ser Ser Pro Gly Ser Trp Ala Cys Gly Leu
                 200                  205                 210
Ala Gly Arg Gly Arg Ser Gly Asp Val Cys Pro Leu Ala Pro Arg
                 215                  220                 225
Ala Val Ala Met Gly Val Arg Ala Gly Ile Pro Ala Trp Gly Gly
```

230 235 240
Arg Ser Arg

<210> 12
<211> 463
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526258CD1

<400> 12
Met Arg Arg Pro Arg Gly Glu Pro Gly Pro Arg Ala Pro Arg Pro
1 5 10 15
Thr Glu Gly Ala Thr Cys Ala Gly Pro Gly Glu Ser Trp Ser Pro
20 25 30
Ser Pro Asn Ser Met Leu Arg Val Leu Leu Ser Ala Gln Thr Ser
35 40 45
Pro Ala Arg Leu Ser Gly Leu Leu Leu Ile Pro Pro Val Gln Pro
50 55 60
Cys Cys Leu Gly Pro Ser Lys Trp Gly Asp Arg Pro Val Gly Gly
65 70 75
Gly Pro Ser Ala Gly Pro Val Gln Gly Leu Gln Arg Leu Leu Glu
80 85 90
Gln Ala Lys Ser Pro Gly Glu Leu Leu Arg Trp Leu Gly Gln Asn
95 100 105
Pro Ser Lys Val Arg Ala His His Tyr Ser Val Ala Leu Arg Arg
110 115 120
Leu Gly Gln Leu Leu Gly Ser Arg Pro Arg Pro Pro Pro Val Glu
125 130 135
Gln Val Thr Leu Gln Asp Leu Ser Gln Leu Ile Ile Arg Asn Cys
140 145 150
Pro Ser Phe Asp Ile His Thr Ile His Val Cys Leu His Leu Ala
155 160 165
Val Leu Leu Gly Phe Pro Ser Asp Gly Pro Leu Val Cys Ala Leu
170 175 180
Glu Gln Glu Arg Arg Leu Arg Leu Pro Pro Lys Pro Pro Pro Pro
185 190 195
Leu Gln Pro Leu Leu Arg Glu Ala Arg Pro Glu Glu Leu Thr Pro
200 205 210
His Val Met Val Leu Leu Ala Gln His Leu Ala Arg His Arg Leu
215 220 225
Arg Glu Pro Gln Leu Leu Glu Ala Ile Thr His Phe Leu Val Val
230 235 240
Gln Glu Thr Gln Leu Ser Ser Lys Val Val Gln Lys Leu Val Leu
245 250 255
Pro Phe Gly Arg Leu Asn Tyr Leu Pro Leu Glu Gln Gln Phe Met
260 265 270
Pro Cys Leu Glu Arg Ile Leu Ala Arg Glu Ala Gly Val Ala Pro
275 280 285
Leu Ala Thr Val Asn Ile Leu Met Ser Leu Cys Gln Leu Arg Cys
290 295 300
Leu Pro Phe Arg Ala Leu His Phe Val Phe Ser Pro Gly Phe Ile
305 310 315
Asn Tyr Ile Ser Gly Thr Pro His Ala Leu Ile Val Arg Arg Tyr
320 325 330
Leu Ser Leu Leu Asp Thr Ala Val Glu Leu Glu Leu Pro Gly Tyr

```
                    335                     340                     345
Arg Gly Pro Arg Leu Pro Arg Arg Gln Gln Val Pro Ile Phe Pro
                    350                     355                     360
Gln Pro Leu Ile Thr Asp Arg Ala Arg Cys Lys Tyr Ser His Lys
                    365                     370                     375
Asp Ile Val Ala Glu Gly Leu Arg Gln Leu Leu Gly Glu Glu Lys
                    380                     385                     390
Tyr Arg Gln Asp Leu Thr Val Pro Pro Gly Tyr Cys Thr Gly Glu
                    395                     400                     405
Gln Gly Ala Gly Gly Arg Pro Gly Glu Thr Glu Pro Trp Leu Arg
                    410                     415                     420
Pro Pro Ala Leu Leu Pro Ser Arg Leu Pro Ala Val Arg Gln Gln
                    425                     430                     435
Leu Trp Cys Cys Ala Ser Arg Glu Asp Pro Gly Pro Leu Pro Ala
                    440                     445                     450
Ile Pro Thr Lys Val Leu Pro Thr Gly Pro Gly Cys Leu
                    455                     460
```

<210> 13
<211> 184
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526311CD1

<400> 13

```
Met Arg Leu Ala Arg Leu Leu Arg Gly Ala Ala Leu Ala Gly Pro
  1                 5                       10                      15
Gly Pro Gly Leu Arg Ala Ala Gly Phe Ser Arg Ser Phe Ser Ser
                    20                      25                      30
Asp Ser Gly Ser Ser Pro Ala Ser Glu Arg Gly Val Pro Gly Gln
                    35                      40                      45
Val Asp Phe Tyr Ala Arg Phe Ser Pro Ser Pro Leu Ser Met Lys
                    50                      55                      60
Gln Phe Leu Asp Phe Gly Ser Val Asn Ala Cys Glu Lys Thr Ser
                    65                      70                      75
Phe Met Phe Leu Arg Gln Glu Leu Pro Val Arg Leu Ala Asn Ile
                    80                      85                      90
Met Lys Glu Ile Ser Leu Leu Pro Asp Asn Leu Leu Arg Thr Pro
                    95                      100                     105
Ser Val Gln Leu Val Gln Ser Trp Tyr Ile Gln Ser Leu Gln Glu
                    110                     115                     120
Leu Leu Asp Phe Lys Asp Lys Ser Ala Glu Asp Ala Lys Ala Ile
                    125                     130                     135
Tyr Glu Arg Pro Arg Arg Thr Trp Leu Gln Val Ser Ser Leu Cys
                    140                     145                     150
Cys Met Ala Cys Lys Met Ile Phe Ile Val Trp Trp Lys Arg Gln
                    155                     160                     165
Arg Lys Ser Ile Ser Ser Lys Thr His Trp Lys His Lys Ser Lys
                    170                     175                     180
Leu Gln Cys Thr
```

<210> 14
<211> 386
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526315CD1

<400> 14
Met Ser Ser Leu Gly Ala Ser Phe Val Gln Ile Lys Phe Asp Asp
1 5 10 15
Leu Gln Phe Phe Glu Asn Cys Gly Gly Gly Ser Phe Gly Ser Val
20 25 30
Tyr Arg Ala Lys Trp Ile Ser Gln Asp Lys Glu Val Ala Val Lys
35 40 45
Lys Leu Leu Lys Ile Glu Lys Glu Ala Glu Ile Leu Ser Val Leu
50 55 60
Ser His Arg Asn Ile Ile Gln Phe Tyr Gly Val Ile Leu Glu Pro
65 70 75
Pro Asn Tyr Gly Ile Val Thr Glu Tyr Ala Ser Leu Gly Ser Leu
80 85 90
Tyr Asp Tyr Ile Asn Ser Asn Arg Ser Glu Glu Met Asp Met Asp
95 100 105
His Ile Met Thr Trp Ala Thr Asp Val Ala Lys Gly Met His Tyr
110 115 120
Leu His Met Glu Ala Pro Val Lys Val Ile His Arg Asp Leu Lys
125 130 135
Ser Arg Asn Val Val Ile Ala Ala Asp Gly Val Leu Lys Ile Cys
140 145 150
Asp Phe Gly Ala Ser Arg Leu His Asn His Thr Thr His Met Ser
155 160 165
Leu Val Gly Thr Phe Pro Trp Met Ala Pro Glu Val Ile Gln Ser
170 175 180
Leu Pro Val Ser Glu Thr Cys Asp Thr Tyr Ser Tyr Gly Val Val
185 190 195
Leu Trp Glu Met Leu Thr Arg Glu Val Pro Phe Lys Gly Leu Glu
200 205 210
Gly Leu Gln Val Ala Trp Leu Val Val Glu Lys Asn Glu Arg Leu
215 220 225
Lys Lys Leu Glu Arg Asp Leu Ser Phe Lys Glu Gln Glu Leu Lys
230 235 240
Glu Arg Glu Arg Arg Leu Lys Met Trp Glu Gln Lys Leu Thr Glu
245 250 255
Gln Ser Asn Thr Pro Leu Leu Leu Pro Leu Val Ala Arg Met Ser
260 265 270
Glu Glu Ser Tyr Phe Glu Ser Lys Thr Glu Glu Ser Asn Ser Ala
275 280 285
Glu Met Ser Cys Gln Ile Thr Ala Thr Ser Asn Gly Glu Gly His
290 295 300
Gly Met Asn Pro Ser Leu Gln Ala Met Met Leu Met Gly Phe Gly
305 310 315
Asp Ile Phe Ser Met Asn Lys Ala Gly Ala Val Met His Ser Gly
320 325 330
Met Gln Ile Asn Met Gln Ala Lys Gln Asn Ser Ser Lys Thr Thr
335 340 345
Ser Lys Arg Arg Gly Lys Lys Val Asn Met Ala Leu Gly Phe Ser
350 355 360
Asp Phe Asp Leu Ser Glu Gly Asp Asp Asp Asp Asp Asp Asp Gly
365 370 375
Glu Glu Glu Asp Asn Asp Met Asp Asn Ser Glu
380 385

<210> 15

<211> 152
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526442CD1

<400> 15
Met Asp Gln Tyr Cys Ile Leu Gly Arg Ile Gly Glu Gly Ala His
1 5 10 15
Gly Ile Val Phe Lys Ala Lys His Val Glu Thr Gly Glu Ile Val
20 25 30
Ala Leu Lys Lys Val Ala Leu Arg Arg Leu Glu Asp Gly Phe Pro
35 40 45
Asn Gln Ala Leu Arg Glu Ile Lys Ala Leu Gln Glu Met Glu Asp
50 55 60
Asn Gln Tyr Val Val Gln Leu Lys Ala Val Phe Pro His Gly Gly
65 70 75
Gly Phe Val Leu Ala Phe Glu Phe Met Leu Ser Asp Leu Ala Glu
80 85 90
Val Val Arg His Ala Gln Arg Pro Leu Ala Gln Ala Gln Val Lys
95 100 105
Ser Tyr Leu Gln Met Leu Leu Lys Gly Val Ala Phe Cys His Ala
110 115 120
Asn Asn Ile Val His Arg Asp Leu Pro Pro Arg Pro Ile Gln Gly
125 130 135
Pro Pro Thr Ser Met Thr Ser Thr Trp Thr Gly Leu Leu Arg Ser
140 145 150
Arg Cys

<210> 16
<211> 4430
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526185CB1

<400> 16
ccggctccag cggccagcgc gcgcgggccc aggccgcccg gctccagccc agcagtagcg 60
gcagcagcgg cggcggcggc agtgcgcgcg aggccctgcg cccccagcag ctcctccctg 120
gcgccgtgca tggagacgcg gcccgccacc cgccgctgag cccccgccgc ccggccggga 180
cccgccaggg ctggggtggc ctcgggctcc ggccggcccc gccgcccgag ggctgcgcgc 240
ggcccgcggg cctcgccgcc ccgcgcggat cgtcgcggcc cggccgtccc gtcccaggaa 300
gtggccgtcc tgagcgccat ggctcactcc ccggtgcagt cgggcctgcc cggcatgcag 360
aacctaaagg cagacccaga agagcttttt acaaaactag agaaaattgg gaagggctcc 420
tttggagagg tgttcaaagg cattgacaat cggactcaga aagtggttgc cataaagatc 480
attgatctgg aagaagctga agatgagata gaggacattc aacaagaaat cacagtgctg 540
agtcagtgtg acagtccata tgtaaccaaa tattatggat cctatctgaa ggatacaaaa 600
ttatggataa taatggaata tcttggtgga ggctccgcac tagatctatt agaacctggc 660
ccattagatg aaacccagat cgctactata ttaagagaaa tactgaaagg actcgattat 720
ctccattcgg agaagaaaat ccacagagac attaaaggca gacatctggt ccctggcat 780
aacagctatt gaacttgcaa gaggggaacc acctcattcc gagctgcacc ccatgaaagt 840
tttattcctc attccaaaga acaacccacc gacgttggaa ggaaactaca gtaaacccct 900
caaggagttt gtggaggcct gtttgaataa ggagccgagc tttagaccca ctgctaagga 960
gttattgaag cacaagttta tactacgcaa tgcaaagaaa acttcctact tgaccgagct 1020

catcgacagg tacaagagat ggaaggccga gcagagccat gacgactcga gctccgagga 1080
ttccgacgcg gaaacagatg gccaagcctc gggggcagt gattctgggg actggatctt 1140
cacaatccga gaaaaagatc ccaagaatct cgagaatgga gctcttcagc catcggactt 1200
ggacagaaat aagatgaaag acatcccaaa gaggcctttc tctcagtgtt tatctacaat 1260
tatttctcct ctgtttgcag agttgaagga gaagagccag gcgtgcggag ggaacttggg 1320
gtccattgaa gagctgcgag gggccatcta cctagcggag gaggcgtgcc ctggcatctc 1380
cgacaccatg gtggcccagc tcgtgcagcg gctccagaga tactctctaa gtggtggagg 1440
aacttcatcc cactgaaatt cctttggcat ttggggtttt gtttttcctt ttttccttct 1500
tcatcctcct cctttttttaa aagtcaacga gagccttcgc tgactccacc gaagaggtgc 1560
gccactggga gccaccccag tgccaggcgc ccgtccaggg acacacacag tcttcactgt 1620
gctgcagcca gatgaagtct ctcagatggg tggggagggt cagctccttc cagcgatcat 1680
tttattttat tttattactt ttgtttttaa ttttaaccat agtgcacata ttccaggaaa 1740
gtgtctttaa aaacaaaaac aaaccctgaa atgtatattt gggattatga taaggcaact 1800
aaagacatga aacctcaggt atcctgcttt aagttgataa ctccctctgg gagctggaga 1860
atcgctctgg tggatgggtg tacagatttg tatataatgt catttttacg gaaacccttt 1920
cggcgtgcat aaggaatcac tgtgtacaaa ctggccaagt gcttctgtag ataacgtcag 1980
tggagtaaat attcgacagg ccataacttg agtctattgc cttgccttta ttacatgtac 2040
attttgaatt ctgtgaccag tgatttgggt tttattttgt atttgcaggg tttgtcatta 2100
ataattaatg cccctctctt acagaacact cctatttgta cctcaacaaa tgcaaatttt 2160
ccccgtttgc cctacgcccc ttttggtaca cctagaggtt gatttccttt ttcatcgatg 2220
gtactatttc ttagtgtttt aaattggaac atatcttgcc tcatgaagct ttaaattata 2280
attttcagtt tctccccatg aagcgctctc gtctgacatt tgtttggaat cgtgccactg 2340
ctggtctgcg ccagatgtac cgtcctttcc aatacgattt tctgttgcac cttgtagtgg 2400
attctgcata tcatctttcc cacctaaaaa tgtctgaatg cttacacaaa taaattttat 2460
aacacgctta ttttgcatac tccttgaaat gtgactcttc agaggacagg gcacctgctg 2520
tgtatgtgtg gccgtgcgtg tgtactcgtg gctgtgtgtg tgtgatgaga cactttggaa 2580
gactccaggg agaagtcccc aggcctggag ctgccgagtg cccaggtcag cgccctggac 2640
tgcttgcgca cttgctcacc gagatgatgc agttggaggt tgctgatctg tgcgattgct 2700
gtagcggttg ccggggacct taagagttat tttgcttctc tggaaggggc ctatgcttgc 2760
taggcaggca gccagtgtgt ctgtttttct tggtttgctg tgggaccttg cttggcgagg 2820
gggaaaatct ctgggtttct ggagtgggag ggttcgtgca gcagctgttg actggtacat 2880
gaagcattct tttatgtttg ttgaagctga tgattgacat ctcccgtggg tgtgccagtt 2940
cttgtggagt taagacagga tttttggaag caaggaagtt agtgggtgag cttggggatg 3000
tagctcagct atctgctggt ctagtggcct ctaagctata gggaggggac agagccctga 3060
gctacagatg cttgagtggg ttattgtgtc ggtttgctag tgcagtctgg tttttaagct 3120
ctaaaattga ggtattttat tagaagtgga tttgggttga actcttaatt tgtataaggg 3180
atatattttg gttggggaaa tagaactgag ttgctaattc ttattgtact cattactcca 3240
tacaagaatg ttatgttgaa taataaaatt ggagaagatt tcattttgtg tttccaggga 3300
gtattctgtg tggggaactg tttccttacg tgaggccggc ggcataagtc aaagatgagt 3360
tttgtccttg cgaatcacac agattgagtc tgtgttcccc agggtgtgcc gttacctgat 3420
ttttaagtga gccagggcgg acagcagctt ttctgattta cagagttctt cagatttaca 3480
aatggacaat gacatcacag tttttagcac tgaagccagt ctcatgctag taacagtggg 3540
tgagccgctc gagggactgg gttctaatga atactggtat gaacggggag tctctgcagt 3600
cgccagacaa atcatactca gccccttccc ccgtagagca acaagtggtt cttttagagt 3660
tgactggcag catttcctgt cgggggaggt ggggtttgat ggagttagaa agctcgcctc 3720
tgtgtacatt ctctcctggg ctgttacttt ctgtagacgc acaaaatcag ccccaatgtt 3780
tttaagggca tcttagccaa ggaagctggc ttttgtgtcg ccacttccag gcctgcatta 3840
agagagagcc caggcaccag ggctaccact ggaacctgcc tcagcgtcaa ctgctgctgg 3900
tctgtagcca ggcccagcct ttgagacggg tttactgtca ccagtagcct ctcagtgcca 3960
gccctgagct gctcctggct cagctgccca gagcctgcag cctggggagg tactcagcct 4020
ctgggagacg agggccgtgg actgggtggc tggtagctcc tgcgtttttg agctgtgtcc 4080
tggctggctg ctgccaatga ggtggacacc agtgtggttt ggggtgcact ggccacttct 4140
tgctgggttc tgattttctt ggaagtgcat ctgccttcct tatccaatag ttttatccct 4200
gcattgctct tgtgaagtgg ctggtttggt tctgtatgta gcattttgta cctttcctct 4260
ggcaaaacac tgtcagttta taaacatttt ttatatttcc ctcctttaaa aacagcttgt 4320
gtatttctgc tataaaatgt gtcagcaaag gcagagtgac ctaatagggc atgttcttaa 4380
gcacagggac tgtatcatgc aggggccaat aaagctcaag aaaacgagta 4430

<210> 17

<211> 3276
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526192CB1

<400> 17
caaaacgccg tggccgtcgc gcggcgccat ccgttgtcgc aaagcggcgc gagaaacgcc 60
cagccgggtg ttggccccgc cccgcgctgt gacgtcggcg gcgcgcgccc ccggcgccgt 120
ggccgcggct gcgcagtggg gcgcgtatgg ctgccgcccc ccgctggcag acgctggcgg 180
cgtaaggcgc gcgggccccg gagcgggcgc ggcggagcgc ggcgagcccg gcgcctcccg 240
tcccgaacat gcggaggccg gcccaggcgg cgcgggagcc ggagcggggg cccaagcggc 300
accggagccg gagcgcgagg gggcgcgggg cccggagcgg gggtccgcgc tgcgctgctg 360
aggccgggcc ggccgcccag acgctgcccg cgggcccggc cacggcggag ccaagctgtg 420
agccgtgagc tttgaggcgg tgggatgtgt cagcagaatg tctcctgccc ccgagagcga 480
ccccgaggcc actgagaaga gcagcgcggc ctggccggcc cgaacgcctg cgtctcagta 540
gctgggagcc acgggcccac gcccgcccac cggccgcagt gatgttctag ccacagagga 600
gccaagacct caggtttcca gagacttggg atttgcacgg cagcagagtc accgtggaga 660
ggccagggta tcacaaactt atggattttg acaagaaagg agggaaaggg gagacggagg 720
agggccggag aatgtccaag gccggcgggg gccggagcag ccacggcatc cggagctcgg 780
ggaccagctc ggggtcctg atggtggcc ccaacttccg cgtcggcaag aagatcggct 840
gcggcaactt cggggagctc cgcctaggaa agaatctcta tacaaatgaa tacgtggcta 900
tcaaattggt gagtcggccc ctccacccca cccccgctga cgtgcccccc agggatttca 960
gggcagcgac ccggtcccct ggtgactcgc tcttgtgccc ccaggagccg atcaagtccc 1020
gggcccccgca gctgcacctg gagtaccggt tctacaagca gctcagcgcc acagagggcg 1080
tccctcaggt ctactacttc ggtccgtgcg ggaagtacaa cgccatggtg ctggagctgc 1140
tggggcccat cctggaggac ctgttcgacc tgtgcgaccg gaccttcacg ctcacgacgg 1200
tgctgatgat cgccatccag ctgatcacgc gcatggagta tgtgcacacc aagagcctaa 1260
tctaccggga cgtgaagccc gagaacttcc tggtgggccg cccggggacc aagcggcagc 1320
atgccatcca catcatcgac ttcgggctgg ccaaggagta catcgacccc gagaccaaga 1380
agcacatccc gtaccgcgag cacaagagcc tgacgggcac ggcgcgctac atgagcatca 1440
acacgcacct gggcaaggag cagagccgcc gcgacgacct ggaggcgctg ggccacatgt 1500
tcatgtactt cctgcgcggc agcctcccct ggcaggggct caaggtgggc gaggaggccg 1560
ggcaggcggg cggggacgca gggcgggagc aaggctgacc acagaccccc gcaggccgac 1620
acgctcaagg agcggtacca gaagatcggg gacaccaaac gcgccacgcc catcgaggtg 1680
ctctgcgaga acttcccaga ggagatggcc acgtacctgc gctatgtgcg gcgcctggac 1740
ttcttcgaga agcccgacta tgactacctg cggaagctct tcaccgacct cttcgaccgc 1800
agtggcttcg tgttcgacta tgagtacgac tgggccggga agcccctgcc gacccccatc 1860
ggcaccgtcc acaccgacct gccctcccag cctcagctcc gggacaaagc ccagccgcac 1920
agcaaaaacc aggcgttgaa ctccaccaac ggggagctga atgcggacga ccccacggcc 1980
ggccactcca acgccccgat cacagcgcct gcagaggtgg aggtggccga tgaaaccaaa 2040
atgctgcacc aaagctcggg cgccgcgggc acggctgctg cagtctcttc ccagcctggc 2100
cctggcaagg ggcgggtggg cgctgccagg cgggtgcttc tcgacgcact tgctcccgga 2160
ggctgcgccc cggcgcctgg aacccgaggt gggaggaccg gttggtgtca ccctgctcgg 2220
ccctcagccc tgccgcgtgg ggcgcgtggg cacggagctt cttgcctctc tgctccgaca 2280
cccggcaagc agccggagac aaaacgcctt aaagccccccg gcccagcccct gcaggtatat 2340
tgcaggggcc tgggggcggc cctggactgg cgggcggttc cccagtgggg tgccctggag 2400
gctgccgggc agagtggagc agcttggggc cgtgcccagg gcggtggctg tgagtctagt 2460
ttttgcttta ccaagtgtac agaaatggca tttacgtttc tctgatgctc ccttgaagcc 2520
atagaattta ggggcttttt taaaaaaata aaagaaaaat gaaaccaaac ccaagtgtag 2580
agggatttgt ctgggctttc cacgaagctt gacctggaac gggcgttgct tccatcccca 2640
tcctgcctgt ccgggacgag tccggagcgg ctggcggcct ccggtaacag aaaccgactg 2700
atgaggcgga aggtaaggaa gatggaagca gagggcagag ctgggctctg tctggggaga 2760
gggcaggaga cgagtgttca cgtaccatgg aaaggggaag tcacacacat gcgacttggc 2820
cccgggggtc ccgttccccg acactacaca aacatacctg aaagcctcag cgacggggcc 2880
caggcaggat ggtcctggct gctctgacgg cggaaggcct ccttgactcc ctctgttcac 2940
gcagcagggc agaaaacatc tccacggggg ccacgacact gtgaagggaa tcagcagtag 3000

ctcccagaag aacagcggaa actgcaggca ggtgaagacc ttgcagcact agccccggct 3060
ccgccccgtg ccttctcccc agacaacacc ccatacccgg cagcaagggt ggaagaccag 3120
taccaccgta atatgttgtg acaaagcaga aataatgcac ctgtaagagt cagatggcaa 3180
gagggaaatg gaatgagctc atcgatggtt ttcccggcag tagcttgggg ataaggacta 3240
cttgtcatgt gctttatata tttacccaca tgttaa 3276

<210> 18
<211> 3910
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526193CB1

<400> 18
tgcgggggc tgggggggga acttagttgt tggcagtttc ttcacgggat gtgtttaaat 60
tgccgagtcc ccacatacgc gccaccccac aaatctcctt cgaggccgtg gaggccacac 120
ggctgccgcc tcgccctctc ctccaggagt atgctgggat ttgtagtcca gcagccggac 180
tgtgccgagc tacctttccc agcttgccct gcggctcggg tgatatcaac agtcttttcc 240
agaactctgt ctgcactgag accctcttcc cccagtcctc ttctcgcggt cgactccttc 300
ccatccgtgg cgacagaacg gcggttgcag gagaggcccc ggtccctcgc cgcgccgccc 360
cgaggggcac ttccggcggc ggttcacttc ctggttgggt ggatggagcc gggcgggagc 420
gcgcgcgggg gaggggcggc gggtcagtct ccgcccggcg ctcccgggat cagctggcgg 480
gcgggcggga gccgagcgcg gccccggctc tcgctgcagc gccgcctctt ctctgcgtcg 540
caggccggcc cggcggccgt gacaatgtcg cggggctggt agcagggcgc cggccgccga 600
gccgtctcaa gtttaaactt acacgaatcg ctttctggag gaggagggga cccgctgcgc 660
gattgacacg catattccta taggcatcct ccctcagccc ccaccccac ggccggattc 720
gggtggctcc tctccgaggt gaaatctgag aagaaatcct tggatctctt ttcttaaaaa 780
aaaaaaaaaa aaaaaaaaaa tctagaaacc atcggtattt tgctttgctg ctccctattc 840
gcaagatgaa gaagtttttc gactcccggc gagagcaggg cggctctggc ctgggctccg 900
gctccagcgg aggaggggc agcacctcgg gcctgggcag tggctacatc ggaagagtct 960
tcggcatcgg gcgacagcag gtcacagtgg acgaggtgtt ggcggaaggt ggatttgcta 1020
ttgtatttct ggtgaggaca agcaatggga tgaaatgtgc cttgaaacgc atgtttgtca 1080
acaatgagca tgatctccag gtgtgcaaga gagaaatcca gataatgagg gatctttcag 1140
ggcacaagaa tattgtgggt tacattgatt ctagtatcaa caacgtgagt agcggtgatg 1200
tatgggaagt gctcattctg atggactttt gtagaggtgg ccaggtggta aacctgatga 1260
accagcgcct gcaaacaggc tttacagaga atgaagtgct ccagatattt tgtgatacct 1320
gtgaagctgt tgcccgcctg catcagtgca aaactcctat tatccaccgg gacctgaagg 1380
ttgaaaacat cctcttgcat gaccgaggcc actatgtcct gtgtgacttt ggaagcgcca 1440
ccaacaaatt ccagaatcca caaactgagg gagtcaatgc agtagaagat gagattaaga 1500
aatacacaac gctgtcctat cgagcaccag aaatggtcaa cctgtacagt ggcaaaatca 1560
tcactacgaa ggcagacatt tgggctcttg gatgtttgtt gtataaatta tgctacttca 1620
ctttgccatt tggggaaagt caggtggcaa tttgtgatgg aaacttcaca attcctgata 1680
attctcgata ttctcaagac atgcactgcc taattaggta tatgttggaa ccagaccctg 1740
acaaaaggcc ggatatttac caggtgtcct acttctcatt taagctactc aagaaagagt 1800
gcccaattcc aaatgtacag aactctccca ttcctgcaaa gcttcctgaa ccagtgaaag 1860
ccagtgaggc agctgcaaaa aagacccagc caaaggccag actgacagat cccattccca 1920
ccacagagac ttcaattgca ccccgccaga ggcctaaagc tgggcagact cagccgaacc 1980
caggaatcct tcccatccag ccagcgctga caccccggaa gagggccact gttcagcccc 2040
cacctcaggc tgcaggatcc agcaatcagc ctggcctttt agccagtgtt ccccaaccaa 2100
aaccccaagc cccacccagc cagcctctgc cgcaaactca ggccaagcag ccacaggctc 2160
ctcccactcc acagcagacg ccttctactc aggcccaggg tctgcccgct caggcccagg 2220
ccacacccca gcaccagcag caactcttcc tcaagcagca acagcagcag caacagccac 2280
cgccagcaca gcagcagccg gcaggcacgt tttaccagca gcagcaggcc cagactcagc 2340
agtttcaggc agtacatcca gcaacccagc aaccagcaat tgctcagttc cctgtggtgt 2400
cccaaggagg ctctcaacag cagctaatgc agaatttcta ccagcagcag cagcagcagc 2460
aacaacaaca gcaacagcaa cagctggcca cagccctgca tcaacaacag ctgatgactc 2520
agcaggctgc cttgcagcaa aagcccacta tggcagcagg acagcagccc cagccacagc 2580

cagctgcagc cccacagcca gcccctgccc aggagccagc gcagattcaa gccccagtaa 2640
gacaacagcc aaaggttcag acaaccccac ctcctgccgt ccaggggcag aaagttggat 2700
ctctcactcc accctcatcc cccaaaaccc aacgtgctgg gcacaggcgt attctcagtg 2760
acgtaaccca cagtgcagtc tttggggtcc ctgccagcaa atcaacccag ctgctccagg 2820
cagctgcagc tgaggccagt ctcaataagt ccaagtctgc aaccaccact ccatcaggct 2880
ctcctcggac ctctcaacaa aacgtttata atccttcaga agggtctacg tggaatccct 2940
ttgatgacga taatttctcc aaactcacag ctgaagaact gctaaacaag gactttgcca 3000
agcttgggga aggcaaacat cccgagaagc ttggaggctc agctgagagt ttgatcccag 3060
gctttcaatc aacccaaggt gatgcttttg ctacgacctc attttctgct ggaactgaaa 3120
aactaattga gggactcaaa tctcctgaca cttctcttct gctccctgac ctcttgccta 3180
tgacagatcc ttttggtagc acttctgatg ctgtaattga aaaagctgat gttgctgttg 3240
agagtctcat accaggactg gagcccccag ttccccagcg cctcccatct cagacggaat 3300
ctgtgacctc gaatcgcaca gattctctca ccggggaaga ttccctgctt gattgctctc 3360
tgctctctaa ccctactact gaccttctgg aagagtttgc ccccacagca atctctgctc 3420
cagtccataa agctgcagaa gatagtaatc tcatctcagg ttttgatgtc cctgagggct 3480
cggacaaggt ggctgaagat gagtttgacc ctattcctgt attgataacc aaaaacccac 3540
aaggtgggca ctctagaaac agcagtggga gctctgagtc cagtcttccc aacctagcca 3600
ggtctttact gctggtggat cagctcatag acctgtagcc gtgacccagt agcagatgca 3660
gttctgtaac cttcataccg taaaatacat tttcattacg gagttatgaa aaaaatgatt 3720
tttttaaaaa aatctgcaaa taagggggccc tccagccctt ttctcctacc ccttgccttc 3780
tcctgtagaa atgataagga aagaaaatca ctttgggcct ccagatattc cttgggcagt 3840
tcctccttgt tagtttgctg tgttttctca ttacccttct tcaatagcat tatcttaaat 3900
caagcactag 3910

<210> 19
<211> 4380
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526196CB1

<400> 19
gccgcggcgg aggggacggg gctaggccgg gtcgccgcct gacgcgacgc gtcctcacgg 60
gcgcctacgt cacggcgtcg aggcggaaga tggtgcacct ccgggccggc ggttgctgag 120
ctgacccgga cggcgaggga gcgggagccc gagcccgacc actccggctg ccgcggggtg 180
cggcgcagcc accgccatgt cgctgctgca gtcggcgctc gacttcttgg cgggtccagg 240
ctccctgggc ggtgcttccg gccgcgacca gagtgacttc gtggggcaga cggtggaact 300
gggcgagctg cggctgcggg tgcggcgggt cctggccgaa ggagggtttg catttgtgta 360
tgaagctcaa gatgtgggga gtggcagaga gtatgcatta aagaggctat tatccaatga 420
agaggaaaag aacagagcca tcattcaaga agtttgcttc atgctctgtt cactcggaga 480
gccccgccggc tgcctgagtg tgggttcggg tggacacagc cacgcctcag cctccctgcg 540
cacagccccc tgagggccct gcctcctcct gccacgcgcg ggatggactt tggtgtcgct 600
gtggtcagtg cacagaactg tggacatggt tatgtacgtt ctcctttaaa caagacaact 660
gcagaaaaag ctttccggcc acccgaacat tgtccagttt tgttctgcag cgtctatagg 720
aaaagaggag tcagacacgg ggcaggctga gttcctcttg ctcacagagc tctgtaaagg 780
gcagctggtg gaatttttga agaaaatgga atctcgaggc ccccttcgt gcgacacggt 840
tctgaagatc ttctaccaga cgtgccgcgc cgtgcagcac atgcaccggc agaagccgcc 900
catcatccac agggacctca aggttgagaa cttgttgctt agtaaccaag ggaccattaa 960
gctgtgtgac tttggcagtg ccacgaccat ctcgcactac cctgactaca gctggagcgc 1020
ccagaggcga gccctggtgg aggaagagat cacgaggaat acaacaccaa tgtatagaac 1080
accagaaatc atagacttgt attccaactt cccgatcggc gagaagcagg atatctgggc 1140
cctgggctgc atcttgtacc tgctgtgctt ccggcagcac ccttttgagg atggagcgaa 1200
acttcgaata gtcaatggga agtactcgat ccccccgcac gacacgcagt acacggtctt 1260
ccacagcctc atccgcgcca tgctgcaggt gaacccggag gagcggctgt ccatcgccga 1320
ggtggtgcac cagctgcagg agatcgcggc cgcccgcaac gtgaacccca agtctcccat 1380
cacagagctc ctggagcaga atggaggcta cgggagcgcc acactgtccc gagggccacc 1440
ccctcccgtg ggccccgctg gcagtggcta cagtggaggc ctggcgctgg cggagtacga 1500

```
ccagccgtat ggcggcttcc tggacattct gcggggtggg acagagcggc tcttcaccaa 1560
cctcaaggac acctcctcca aggtcatcca gtccgtcgct aattatgcaa agggtgacct 1620
ggacatatct tacatcacat ccagaattgc agtgatgtca ttcccagcag aaggtgtgga 1680
gtcagcgctc aaaaacaaca tcgaagatgt gcggttgttc ctggactcca agcacccagg 1740
gcactatgcc gtctacaacc tgtccccgag gacctaccgg ccctccaggt tccacaaccg 1800
ggtctccgag tgtggctggg cagcacggcg ggccccacac ctgcacaccc tgtacaacat 1860
ctgcaggaac atgcacgcct ggctgcggca ggaccacaag aacgtctgcg tcgtgcactg 1920
catggacggg agagccgcgt ctgctgtggc cgtctgctcc ttcctgtgct tctgccgtct 1980
cttcagcacc gcggaggccg ccgtgtacat gttcagcatg aagcgctgcc caccaggcat 2040
ctggccatcc cacaaaaggt acatcgagta catgtgtgac atggtggcgg aggagcccat 2100
cacaccccac agcaagccca tcctggtgag ggccgtggtc attgacaccc gtgccgcgtg 2160
ttcagcaagc agaggagcgg ctgcaggccc ttctgcgagg tctacgtggg ggacgagcgt 2220
gtggccagca cctcccagga gtacgacaag atgcgggact ttaagattga agatggcata 2280
ggggtgattc ccctgggcgt cacggtgcaa ggagacgtgc tcatcgtcat ctatcacgcc 2340
cggtccactc tgggcggccg gctgcaggcc aagatggcat ccatgaagat gttccagatt 2400
cagttccaca cggggtttgt gcctcggaac gccaccactg tgaaatttgc caagtatgac 2460
ctggacgcgt gtgacattca agaaaaatac ccggatttat ttcaagtgaa cctggaagtg 2520
gaggtggagc ccagggacag gccgagccgg gaagccccac catgggagaa ctcgagcatg 2580
agggggctga accccaaaat cctgttttcc agccgggagg agcagcaaga cattctgtct 2640
aagtttggga agccggagct tccccggcag cctggctcca cggctcagta tgatgctggg 2700
gcagggtccc cggaagccga acccacagac tctgactcac cgccaagcag cagcgcggac 2760
gccagtcgct tcctgcacac gctggactgg caggaagaga aggaggcaga gactggtgca 2820
gaaaatgcct cttccaagga gagcgagtct gccctgatgg aggacagaga cgagagtgag 2880
gtgtcagatg aaggggggatc cccgatctcc agcgagggcc aggaacccag ggccgaccca 2940
gagccccccg gcctggcagc agggctggtg cagcaggact tggtttttga ggtggagaca 3000
ccggctgtgc tgccagagcc tgtgccacag gaagacgggg tcgacctcct gggcctgcac 3060
tccgaggtgg gcgcagggcc agctgtaccc ccgcaggcct gcaaggcccc ctccagcaac 3120
accgacctgc tcagctgcct ccttgggccc cctgaggccg cctcccaggg gcccccggag 3180
gatctgctca gcgaggaccc gctgctcctg gcaagcccgg cccctcccct gagcgtgcag 3240
agcccccaag gaggaggccc cctgccgctg ctgacccctt tggcccgctt ctgccgtctt 3300
caggcaacaa ctcccagccc tgctccaatc ctgatctctt cggcgaattt ctcaattcgg 3360
actctgtgac cgtcccacca tccttcccgt ctgcccacag cgctccgccc ccatcctgca 3420
gcgccgactt cctgcacctg ggggatctgc caggagagcc cagcaagatg acagcctcgt 3480
ccagcaaccc agacctgctg ggaggatggg ctgcctggac cgagactgca gcgtcggcag 3540
tggcccccac gccagccaca gaaggccccc tcttctctcc tggaggtcag ccggcccctt 3600
gtggctctca ggccagctgg accaagtctc agaacccgga cccatttgct gaccttggcg 3660
acctcagctc cggcctccaa gacccccaag cccagagcac agtgagccca aggggacagc 3720
gtgtctgcac ctgttccagg cgactgccaa ctggcaagct aaaaccggga gttgctgaca 3780
ctggcactgc tgccagcccc caccggcatt gtggctcacc agctggattc cctcctgggg 3840
gcttcattcc caaaacggcc accacgccca aaggcagcag ctcctggcag acaagtcggc 3900
cgccagccca gggcgcctca tggccccctc aggccaagcc gccccccaaa gcctgcacac 3960
agccaaggcc taactatgcc tcgaacttca gtgtgatcgg ggcgcgggaa gagcgggggg 4020
tccgcgcacc agctttgctc aaaaccaaaa gtctctgaga acgactttga agattgttgt 4080
ccaataaggc ttctcctcca ggtctgaaag aaagggcaaa gacattgcag agatgaggag 4140
aggacctggc taaagacacg gacccactca agctgaagct cctggactgg attgagggca 4200
aggagcggaa catccgggcc ctgctgtcca cgctgcacac agtgctgtgg gacggggaga 4260
gccgctggac gcccgtgggc atggccgacc tggtggctcc ggagcaagtg aagaagcact 4320
atcgccgcgc ggtgctggct gtgcaccccg acaaggtgag cagagctgcc aggcggccgc 4380
```

<210> 20
<211> 4293
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526198CB1

<400> 20

gccgcggcgg aggggacggg gctaggccgg gtcgccgcct gacgcgacgc gtcctcacgg 60
gcgcttacgt cacggcgtcg aggcggaaga tggtgcacct ccgggccggc ggttgctgag 120
ctgacccgga cggcgaggga gcgggagccc gagcccgacc actccggctg ccgcggggtg 180
cggcgcagcc accgccatgt cgctgctgca gtcggcgctc gacttcttgg cgggtccagg 240
ctccctgggc ggtgcttccg gccgcgacca gagtgacttc gtggggcaga cggtggaact 300
gggcgagctg cggctgcggg tgcggcgggt cctggccgaa ggagggtttg catttgtgta 360
tgaagctcaa gatgtgggga gtggcagaga gtatgcatta aagaggctat tatccaatga 420
agaggaaaag aacagagcca tcattcaaga agtttgcttc atgaaaaagc tttccggcca 480
cccgaacatt gtccagtttt gttctgcagc gtctatagga aaagaggagt cagacacggg 540
gcaggctgag ttcctcttgc tcacagagct ctgtaaaggg cagctggtgg aatttttgaa 600
gaaaatggaa tctcgaggcc cccttttcgtg cgacacggtt ctgaagatct tctaccagac 660
gtgccgcgcc gtgcagcaca tgcaccggca gaagccgccc atcatccaca gggacctcaa 720
ggttgagaac ttgttgctta gtaaccaagg gaccattaag ctgtgtgact ttggcagtgc 780
cacgaccatc tcgcactacc ctgactacag ctggagcgcc cagaggcgag ccctggtgga 840
ggaagagatc acgaggaata caacaccaat gtatagaaca ccagaaatca tagacttgta 900
ttccaacttc ccgatcggcg agaagcagga tatctgggcc ctgggctgca tcttgtacct 960
gctgtgcttc cggcagcacc cttttgagga tggagcgaaa cttcgaatag tcaatgggaa 1020
gtactcgatc cccccgcacg acacgcagta cacggtcttc cacagcctca tccgcgccat 1080
gctgcaggtg aacccggagg agcggctgtc catcgccgag gtggtgcacc agctgcagga 1140
gatcgcggcc gcccgcaacg tgaaccccaa gtctcccatc acagagctcc tggagcagaa 1200
tggaggctac gggagcgcca cactgtcccg agggccaccc cctcccgtgg gccccgctgg 1260
cagtggctac agtggaggcc tggcgctggc ggagtacgac cagccgtatg gcggcttcct 1320
ggacattctg cggggtggga cagagcggct cttcaccaac ctcaaggaca cctcctccaa 1380
ggtcatccag tccgtcgcta attatgcaaa gggtgacctg gacatatctt acatcacatc 1440
cagaattgca gtgatgtcat tcccagcaga aggtgtggag tcagcgctca aaaacaacat 1500
cgaagatgtg cggttgttcc tggactccaa gcacccaggg cactatgccg tctacaacct 1560
gtccccgagg acctaccggc cctccaggtt ccacaaccgg gtctccgagt gtggctgggc 1620
agcacggcgg gccccacacc tgcacaccct gtacaacatc tgcaggaaca tgcacgcctg 1680
gctgcggcag gaccacaaga acgtctgcgt cgtgcactgc atggacggga gagccgcgtc 1740
tgctgtggcc gtctgctcct tcctgtgctt ctgccgtctc ttcagcaccg cggaggccgc 1800
cgtgtacatg ttcagcatga agcgctgccc accaggcatc tggccatccc acaaaaggta 1860
catcgagtac atgtgtgaca tggtggcgga ggagcccatc acaccccaca gcaagcccat 1920
cctggtgagg gccgtggtca tgacacccgt gccgctgttc agcaagcaga ggagcggctg 1980
caggcccttc tgcgaggtct acgtggggga cgagcgtgtg gccagcacct cccaggagta 2040
cgacaagatg cgggacttta agattgaaga tggcatagcg gtgattcccc tgggcgtcac 2100
ggtgcaagga gacgtgctca tcgtcatcta tcacgcccgg tccactctgg gcggccggct 2160
gcaggccaag atggcatcca tgaagatgtt ccagattcag ttccacacgg ggtttgtgcc 2220
tcggaacgcc accactgtga aatttgccaa gtatgacctg gacgcgtgtg acattcaaga 2280
aaaatacccg gatttatttc aagtgaacct ggaagtggag gtggagccca gggacaggcc 2340
gagccgggaa gccccaccat gggagaactc gagcatgagg gggctgaacc ccaaaatcct 2400
gttttccagc cgggaggagc agcaagacat tctgtctaag tttgggaagc cggagcttcc 2460
ccggcagcct ggctccacgg ctcagtatga tgctggggca gggtccccgg aagccgaacc 2520
cacagactct gactcaccgc caagcagcag cgcggacgcc agtcgcttcc tgcacacgct 2580
ggactggcag gaagagaagg aggcagagac tggtgcagaa aatgcctctt ccaaggagag 2640
cgagtctgcc ctgatggagg acagagacga gagtgaggtg tcagatgaag gggatcccc 2700
gatctccagc gagggccagg aacccagggc cgacccagag cccccggcc tggcagcagg 2760
gctggtgcag caggacttgg tttttgaggt ggagacaccg gctgtgctgc cagagcctgt 2820
gccacaggaa gacggggtcg acctcctggg cctgcactcc gaggtgggcg cagggccagc 2880
tgtacccccg caggcctgca aggccccctc cagcaacacc gacctgctca gctgcctcct 2940
tgggcccccct gaggccgcct cccaggggcc cccggaggat ctgctcagcg aggacccgct 3000
gctcctggca agcccggccc ctcccctgag cgtgcagagc accccaagag gagggccccc 3060
tgccgctgct gacccctttg gcccgcttct gccgtcttca ggcaacaact cccagccctg 3120
ctccaatcct gatctcttcg gcgaatttct caattcggac tctgtgaccg tcccaccatc 3180
cttcccgtct gcccacagtg ctccgccccc atcctgcagc gccgacttcc tgcacctggg 3240
ggatctgcca ggagagccca gcaagatgac agcctcgtcc agcaacccag acctgctggg 3300
aggatggggct gcctggaccg agactgcagc atcggcagtg gcccccacgc cagccacaga 3360
aggccccctc ttctctcctg gaggtcagcc ggccccttgt ggctctcagg ccagctggac 3420
caagtctcag aacccggacc catttgctga ccttggcgac ctcagctccg gcctccaaga 3480
cccccaagcc cagagcacag tgagcccaag ggacagcgt gtctgcacct gttccaggcg 3540

actgccaact ggcaagctaa aaccgggagt tgctgacact ggcactgctg ccagcccccca 3600
ccggcattgt ggctcaccag ctggattccc tcctgggggc ttcattccca aaacggccac 3660
cacgcccaaa ggcagcagct cctggcagac aagtcggccg ccagcccagg gcgcctcatg 3720
gccccctcag gccaagccgc cccccaaagc ctgcacacag ccaaggccta actatgcctc 3780
gaacttcagt gtgatcgggg cgcgggagga gcgggggtc cgcgcaccca gctttgctca 3840
aaagccaaaa gtctctgaga acgactttga agatctgttg tccaatcaag gcttctcctc 3900
caggtctgac aagaaagggc caaagaccat tgcagagatg aggaagcagg acctggctaa 3960
agacacggac ccactcaagc tgaagctcct ggactggatt gagggcaagg agcggaacat 4020
ccgggccctg ctgtccacgc tgcacacagt gctgtgggac ggggagagcc gctggacgcc 4080
cgtgggcatg gccgacctgg tggctccgga gcaagtgaag aagcactatc gccgcgcggt 4140
gctggccgtg caccccgaca aggctgcggg gcagccgtac gagcagcacg ccaagatgat 4200
cttcatggag ctgaatgacg cctggtcgga gtttgagaac cagggctccc ggcccctctt 4260
ctgaggccgc agtggtggtg gctgcgcaca cag 4293

<210> 21
<211> 6538
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526208CB1

<400> 21
ggagttactc agaagggaag ggaaggtgtg gttgtgcggc ggagtttttg ctttcattct 60
tttaacgttc acagccaaag caaaggcctt tggggattgc cagagtctca gccaccatcc 120
tggaaaacag cgggggaggt gggcctggag gtggcaagtg taatgtggct caggggccgt 180
cattgcccct tgcagaaggg gctgcggggg agggagaaaa cctgcgcccg gttctgggga 240
gctggcgacg cagtgaaccc tgctgaggct gggttttgcc ccgacagtcg ctggtggctg 300
tgggaagggt tgggaccctt ctctgagagc agtgaacagc ccacatccgg cccctgctgt 360
gtcaactctg agcggcgtgg agatgaagtg gttgctctcc cttgctcggc ccaccgggtg 420
tcgtggcccg ggaaccggcc tggagaagtc cctgctgccc ggcgcccaaa acaggggcgt 480
gggcttccgc gacccagggc ggctgccccg ggccatcctc gagttgccct gcatcttccc 540
gctcagtcag ccccagattg aggcagcctt ctctgtgcgg gtttaaatgg gtaactgtga 600
cttctcgcct cattcaccca aacctccagt cttctccccc gcacatcctc ctccacccac 660
ctggtttctc cctagactgg tgtgctcgtg tgtgcaacca aaggagggag tgcgagagat 720
ccacgaaggg acaggcttgg agtcgctaga gggaggtgtg ggaccagcga ggagggggct 780
tcgccaggga gggggtgctg gcaggcggag ggagcggcgg gaggaggcgc cggaggagga 840
gacggaggcc tggggacggc agaagaggct tcgcctgagc cgagcgctct ttctctcgcc 900
gcgccgtctt gaagccgcgc gggctcgtga gcagcgcgag gccgccaagg tgcctcgctt 960
cgcccggagcc gctgccgccc gccggaggga agccggcctc gggcgcgcac gctcgtcgga 1020
gccccggcgc gccccgcgcc tgagcctgct gacagcgccg cggggccagt cccggggtta 1080
gccgcgcgtc tgctcgcttc tggtccgtcg cgctcccagc cagggcacag cccggaccga 1140
ggatggcttc gaccacaacc tgcaccaggt tcacggacga gtatcagctt ttcgaggagc 1200
ttggaaaggg ggcattctca gtggtgagaa gatgtatgaa aattcctact ggacaagaat 1260
atgctgccaa aattatcaac accaaaaagc tttctgctag ggtgcgactt catgatagca 1320
tatcagaaga gggctttcac tacttggtgt ttgatttagt tactggaggt gaactgtttg 1380
aagacatagt ggcaagagaa tactacagtg aagctgatgc cagtcattgc attcagcaga 1440
tcctggaggc tgtgctacac tgccatcaga tgggcgtggt ccatcgggac ctgaagcctg 1500
agaatttgct tttagctagc aaatccaagg gagcagctgt gaaattggca gactttggct 1560
tagccataga agttcaaggg gaccagcagg cgtggtttgg ttttgctggc acacctggat 1620
atctttctcc agaagtttta cgtaaagatc cttatggaaa gccagtggat atgtgggcat 1680
gtggtgtcat tctctatatt ctacttgtgg ggtatccacc cttctgggat gaagaccaac 1740
acagactcta tcagcagatc aaggctggag cttatgattt tccatcacca gaatggggaca 1800
cggtgactcc tgaagccaaa gacctcatca ataaaatgct tactatcaac cctgccaaac 1860
gcatcacagc ctcagaggca ctgaagcacc catggatctg tcaacgttct actgttgctt 1920
ccatgatgca cagacaggag actgtagact gcttgaagaa atttaatgct agaagaaaac 1980
taaagggtgc catcttgaca actatgctgg ctacaaggaa tttctcagca gccaagagtt 2040
tgttgaagaa accagatgga gtaaagaaaa ggaagtccag ttcgagtgtt cagatgatgg 2100

agtcaactga gagttcaaat acaacaattg aggatgaaga tgtggaagca cgaaagcaag 2160
agattatcaa agtcactgaa caactgatcg aagctatcaa caatggggac tttgaagcct 2220
acacaaaaat ctgtgaccca ggccttactg cttttgaacc tgaagctttg ggtaatttag 2280
tggaagggat ggattttcac cgattctact ttgaaaatgc tttgtccaaa agcaataaac 2340
caatccacac tattattcta aaccctcatg tacatctggt aggggatgat gccgcctgca 2400
tagcatatat taggctcaca cagtacatgg atggcagtgg aatgccaaag acaatgcagt 2460
cagaagagac tcgtgtgtgg caccgccggg atggaaagtg gcagaatgtt cattttcatc 2520
gctcggggtc accaacagta cccatcaagc caccctgtat tccaaatggg aaagaaaact 2580
tctcaggagg cacctctttg tggcaaaaca tctaaggcct gaaaaccatt cacatatggg 2640
tcttctaaat ttcaacagtg ccacttctgc attctctgtt ctcaaggcac ctggatggtg 2700
accctgggcc gtcctctcct cctcttcatg catgtttctg agtgcatgaa gttgtgaagg 2760
tcctacatgt aatgcatatg tgatgcatca tcttatcata tattccttcc tatacattgt 2820
ttacacttca actacgggga tgttccacac aaacttaaat tactgttggc aaaacaatag 2880
ggggagatta gacaaaaaaa aaaatccaca atattccaag tacaactctt catcaagttt 2940
ctctgttaat gccaagattt aacagactta agaactattg ttctctgaat gacagttgta 3000
agagaaatgt aaatttttta gaactctttg ctgttaatct gttttggttt gtttggtttt 3060
tttttttttt tttaaggtaa aaaaaaaata caccttcagt ttcctggtgt gatcctggtt 3120
aaaatggatg atttttcatt gaaagttttg ctgattaaca attaaagtgg gatgatatgt 3180
gggcaaaatc acttatgaaa gtagaagcaa gaatcagttg gtttgctacc acataaagcc 3240
atgctgtttt tggtcaaact gtgtaaactg gaaaaattca catcatttct gagtttaatc 3300
actttaggat atattcacat tgttttggtg aatttgctga attgaattgt ttttctttct 3360
caaatctgtg atctcttttc tttatcctgt ttctttgttc ctttcgtttg ctttcttatt 3420
tttcttttgg ttccattctt ttcttacttt tttccctttt ccttttttgg ggaggctggc 3480
tagtagtgtg tgagaaaaga atagaagtga aatttgcata atgaatgtaa aagggaaata 3540
aaagtctttt gaaggtagct atactagcac ttttgatcat cttcagggcc cacaaaaatg 3600
ttgtcaagat tttaaaggtt tataattctg cttaagctct agtttggact taggtatcct 3660
aactatgttg gaggtatttg cattgtttaa agttaggata aaagcaagtt cctcctgtga 3720
ctgcaacgtc ttactgattg ggacagttgc caggaggata ccaacttgat agcagagggg 3780
gttttatgca aacgcactca cctccgcctt ggggaatgaa agggtcactt ctgcatcatc 3840
actagctagt tttctagtgt tagagaggct tacaaatgtt tgccattctc ataagtgttt 3900
tgaacttgat ctttgtgact tgtgcttttt ttagcttctc tcttgaatca gagtatcatt 3960
gtcttcctcc aaggagttag aatttcccag tttaaaacaa aaagggaaat gtcctaggtt 4020
ttctttgtgc ttctcatttt tcctttgttg attcaattcc tgtgattttt gttctcttcc 4080
ctgaagtgct ttacagtgca tggaatctcc atcattgtta ttttaacgat agtaattcac 4140
agtcctcaga agcctatttt taaagcagaa gcaaaaaaga aaaacaaaat aacaaaaaca 4200
acccttcctc ttttctctca tctcacctct ctgtgttgat tactaatcat cttagatatt 4260
attgctagtg gatgtatggt agatgggttg aagcttttct gataattatt acacaattta 4320
aaacaacata tatatttaaa ataaatatat acagtaaata tattgagcca tgttaacctg 4380
ccaatgagat ctgtgaaaaa ataatggcct cattttctc ttttttaattt cttttaccct 4440
tttgtgaagc agctatacgt ggcatacatg tatttaaaga aaaaaaaata gatgtagagt 4500
gtttttttta cacttttaac ttagcatgtg gtgttgaagt attactgtag atcaagtttg 4560
tcttccgcac taagatgtga ggaaattgtg atttgttctc tccaccacaa atgaattaca 4620
catttattat cttctatcat tttgaaacac tgcagtttac catgggacac tgtatatatt 4680
tcttgccata atggtaaagg actgattgat atatttaaga gttaataaat ttgtgatttc 4740
tgctgacagt gcgtccatct ttatttcttc agaagaggta ctgtatgtat gcctgcatag 4800
tgctggccag tgtcaagggc agtgtgtcct actctggtct catttagtac ataacaattt 4860
gcacttggtg aggaggacaa tatagttaac aactaagact cctaaaagct tctctaaact 4920
gtaccctcca atccagcctt cacatggctg cttttttttt tttttttaat acgaacctgt 4980
ccttgtaaca ctttgatgtt atcatttctg ggatacaggc aagcacccca gctcctgcta 5040
ctccccagct tgaacttgag catacatgga tgctcagctt cttttgattt gctaaaaaca 5100
tcacacttgc tcacatgcct gtttatgctg ttcatgttgt ttatgtttct tacctagaat 5160
aaatagtctc ttcccctact tcttttcctg acttcttact ttttcctaag attcagtgta 5220
cagcatcatg ctccacagca aaccttccta ggccctattc tgggcttgcc ttccctctca 5280
aaacctacat aatagattgt atttacctct cctgtcaacc acattgtttt gaaaatatat 5340
ttctatttgt gtctcctcta ctgcagtata atgtctccat gggcaagaac tgtgtattca 5400
tcattgcatt cctaaaccca aaccaaggcc aggaatggag atatcattga taaatagttg 5460
ttgaattgag gccaagccct tttgataaca gaagcctcaa ggggtaccca gatagtcctt 5520
gttttaatga tgggttctct caccactgtc ttgatgctct gagcaagtta cctcttccct 5580
ctgaccctca gtttccatat ttgtaaaatg agaataaaca taccaactta ataaagatat 5640

tgtgaggatt aatgggtaca gagtgactag aatgatattt gatagaaatt aaatggtagc 5700
agtataacta ttctgatcac tgacattaat attcctattg ttattattct ttactcacga 5760
gggtatacaa ctcttgtttt gctgttgggc tgccctcttt atgtaggttt actgttaatg 5820
ctgaggatat actcggactc aaatgtctca gcagaaggct gagagacacc aaatgaagtg 5880
gtcatctagc tgaatgtagg aaaaatgaaa tgtagtagca aatcagtata ttctaaggaa 5940
attttcaagg aatattaatc ttcacccaaa ttttgaattt ttatgtaaaa aattataatt 6000
taagggtaaa catagatgac acagctttcg agtgatttca ttgaataaaa ttctactgac 6060
ttctatgaac ctttcatggc tctgtggtct ttttatcaga tttttaaag gtgagaatgt 6120
acaaaaagat tacaatgaag atcaggtact agaccatgtg tccatgaacg tgaacaaaca 6180
gctgtcataa accaccctaa cctgagaaag cagcaggaag catttacagc attcctgctt 6240
ttctctcaga caaaaccaat tctcagaaga gagctagaat gttctcctgc agactggagt 6300
aggaaaagtt gataacagat taagcagtaa ttgtactcca gaaggatttg catttaggct 6360
tttgctgctt tacaacagaa aaaaaaaatt cttgtttgtc cgtaaaaagt gtttttatgt 6420
ttttttaaat gtcaccaaca tttaaaaatt ggatatgtca tgtaaaagtc aagatttctg 6480
gcttaattaa tttgaaaaag tgtaaggtct gccccactgg ttctgtgttc actacagc 6538

<210> 22
<211> 2349
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526212CB1

<400> 22
ggagttactc agaagggaag ggaaggtgtg gttgtgcggc ggagtttttg ctttcattct 60
tttaacgttc acagccaaag caaaggcctt tggggattgc cagagtctca gccaccatcc 120
tggaaaacag cgggggaggt gggcctggag gtggcaagtg taatgtggct caggggccgt 180
cattgcccct tgcagaaggg gctgcggggg agggagaaaa cctgcgcccg gttctgggga 240
gctggcgacg cagtgaaccc tgctgaggct gggttttgcc ccgacagtcg ctggtggctg 300
tgggaagggt tgggaccctt ctctgagagc agtgaacagc ccacatccgg cccctgctgt 360
gtcaactctg agcggcgtgg agatgaagtg gttgctctcc cttgctcggc ccaccgggtg 420
tcgtggcccg ggaaccggcc tggagaagtc cctgctgccc ggcgcccaaa acaggggcgt 480
gggcttccgc gacccagggc ggctgccccg ggccatcctc gagttgccct gcatcttccc 540
gctcagtcag ccccagattg aggcagcctt ctctgtgcgg gtttaaatgg gtaactgtga 600
cttctcgcct cattcaccca aacctccagt cttctccccc gcacatcctc ctccacccac 660
ctggtttctc cctagactgg tgtgctcgtg tgtgcaacca aaggagggag tgcgagagat 720
ccacgaaggg acaggcttgg agtcgctaga gggaggtgtg ggaccagcga ggagggggct 780
tcgccaggga gggggtgctg gcaggcggag ggagcggcgg gaggaggcgc cggaggagga 840
gacggaggcc tggggacggc agaagaggct tcgcctgagc cgagcgctct ttctctcgcc 900
gcgccgtctt gaagccgcgc gggctcgtga gcagcgcgag gccgccaagg tgcctcgctt 960
cgccggagcc gctgccgccc gccggaggga agccggcctc gggcgcgcac gctcgtcgga 1020
gcccggcgc gccccgcgcc tgagcctgct gacagcgccg cggggccagt cccggggtta 1080
gccgcgcgtc tgctcgcttc tggtccgtcg cgctcccagc cagggcacag cccggaccga 1140
ggatggcttc gaccacaacc tgcaccaggt tcacggacga gtatcagctt ttcgaggagc 1200
ttggaaaggg ggcattctca gtggtgagaa gatgtatgaa aattcctact ggacaagaat 1260
atgctgccaa aattatcaac accaaaaagc tttctgctag ggtgcgactt catgatagca 1320
tatcagaaga gggctttcac tacttggtgg ttgatttagt tactggaggt gaactgtttg 1380
aagacatagt ggcaagagaa tactacagtg aagctgatgc cagtcattgc attcagcaga 1440
tcctggaggc tgtgctacac tgccatcaga tgggcgtggt ccatcgggac ctgaagcctg 1500
agaatttgct tttagctagc aaatccaagg gagcagctgt gaaattggca gactttggct 1560
tagccataga agttcaaggg gaccagcagg cgtggtttgg ttttgctggc acacctggat 1620
atctttctcc agaagtttta cgtaaagatc cttatggaaa gccagtggat atgtgggcat 1680
gtggtgtcat tctctatatt ctacttgtgg ggtatccacc cttctgggat gaagaccaac 1740
acagactcta tcagcagatc aaggctggag cttatgattt tccatcacca gaatggggaca 1800
cggtgactcc tgaagccaaa gacctcatca ataaaatgct tactatcaac cctgccaaac 1860
gcatcacagc ctcagaggca ctgaagcacc catggatctg tcaacgttct actgttgctt 1920
ccatgatgca cagacaggag actgtagact gcttgaagaa atttaatgct agaagaaaac 1980

taaagggtgc catcttgaca actatgctgg ctacaaggaa tttctcagca gccaagagtt 2040
tgttgaagaa accagatgga gtaaaggagt caactgagag ttcaaataca acaattgagg 2100
atgaagatgt gaaaggcacg gtggctcacg cctgtaatcc cagcactttg ggaggtcgag 2160
gcgggcagat cacctgaggt caggagttca agaccagcat ggccaacatg gtgaaaccct 2220
gtctctacta aaaatacaaa aattagctgg gtgtggtggc aggcacctgt aatcccagct 2280
actctggagg ctgagacagg agaatcgctt gaacccggga ggtggaggtt gcagtgagcc 2340
gagatcaca 2349

<210> 23
<211> 8015
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526213CB1

<400> 23
agccgggcag ctgcagcgga gccgcggagc gggcggcggg gcccaggctg tgcgcttggg 60
gagcgcggaa tgtgaggctt ggcgggccgc agcacgctcg gacgggccag gggcggcgac 120
ccctcgcgga cgcccggctg cgcgccgggc cggggacttg cccttgcacg ctccctgcgc 180
cctccagctc gccggcggga ccatgaagaa gttctctcgg atgcccaagt cggagggcgg 240
cagcggcggc ggagcggcgg gtggcggggc tggcggggcc ggggccgggg ccggctgcgg 300
ctccggcggc tcgtccgtgg gggtccgggt gttcgcggtc ggccgccacc aggtcaccct 360
ggaagagtcg ctggccgaag tgatacagat gctgccggtt caggaaccac gtcttgagta 420
ccgagtacca ctgatttcga gcgggcgaag aagactaaga agaaggtgct agagaggtgg 480
attctccaca gttttcctcg tgcgtactca cggtggaatc cgatgtgcat tgaagcgaat 540
gtatgtcaat aacatgccag acctcaatgt ttgtaaaagg gaaattacaa ttatgaaaga 600
gctatctggt cacaaaaata ttgtgggcta tttgggctgt gctgttaatt caattagtga 660
taatgtatgg gaagtcctta tcttaatgga atattgtcga gctggacagg tagtgaatca 720
aatgaataag aagctacaga cgggttttac agaaccagaa gtgttacaga tattctgtga 780
tacctgtgaa gctgttgcaa ggttgcatca gtgtaagact ccaataattc accgggatct 840
gaaggtagaa aatattttgt tgaatgatgg tgggaactat gtactttgtg actttggcag 900
tgccactaat aaatttctta atcctcaaaa agatggagtt aatgtagtag aagaagaaat 960
taaaaagtat acaactctgt catacagagc cctgaaatg atcaaccttt atggagggaa 1020
acccatcacc accaaggctg atatctgggc actgggatgt ctactctata aactttgttt 1080
cttcactctt ccttttggtg agagtcaggt tgctatctgt gatggcaact tcaccatccc 1140
agacaattct cgttactccc gtaacataca ttgcttaata aggttcatgc ttgaaccaga 1200
tccggaacat agacctgata tatttcaagt gtcatatttt gcatttaaat ttgccaaaaa 1260
ggattgtcca gtctccaaca tcaataattc ttctattcct tcagctcttc ctgaaccgat 1320
gactgctagt gaagcagctg ctaggaaaag ccaaataaaa gccagaataa cagataccat 1380
tggaccaaca gaaacctcaa ttgcaccaag acaaagacca aaggccaact ctgctactac 1440
tgccactccc agtgtgctga ccattcaaag ttcagcaaca cctgttaaag tccttgctcc 1500
tggtgaattc ggtaaccata gaccaaaagg ggcactaaga cctggaaatg gccctgaaat 1560
tttattgggt cagggacctc ctcagcagcc gccacagcag catagagtac tccagcaact 1620
acagcaggga gattggagat tacagcaact ccatttacag catcgtcatc ctcaccagca 1680
gcagcagcag cagcagcagc aacagcaaca gcagcagcag caacagcaac agcagcagca 1740
gcagcagcag cagcagcagc agcaccacca ccaccaccac cacactactt caagatgctt 1800
atatgcagca gtatcaacat ggcaacacag cagcaacaga tgcttcaaca acaatttta 1860
atgcattcgg tatatcaacc acaaccttct gcatcacagt atcctacaat gatgccgcag 1920
tatcagcagg ctttctttca acagcagatg ctagctcaac atcagccgtc tcaacaacag 1980
gcatcacctg aatatcttac ctcccctcaa gagttctcac cagccttagt ttcctacact 2040
tcatcacttc cagctcaggt tggaaccata atggactcct cctatagtgc caataggtca 2100
gttgctgata aagaggccat tgcaaatttc acaaatcaga agaacatcag caatccacct 2160
gatatgtcag ggtggaatcc ttttggagag gataatttct ctaagttaac agaagaggaa 2220
ctattggaca gagaatttga ccttctaaga tcaagttctc ctgaaaagaa agctgaacat 2280
tcatctataa atcaagaaaa tggcactgca aaccctatca agaacggtaa aacaagtcca 2340
gcatctaaag atcagcggac tggaaagaaa acctcagtac agggtcaagt gcaaaagggg 2400
aatgatgaat ctgaaagtga ttttgaatca gatcccctt ctcctaagag cagtgaagag 2460

```
gaagagcaag atgatgaaga agttcttcag ggggaacaag gagattttaa tgatgatgat 2520
actgaaccag aaaatctggg tcataggcct ctcctcatgg attctgaaga tgaggaagaa 2580
gaggagaaac atagctctga ttctgattat gagcaggcta aagcaaagta cagtgacatg 2640
agctctgtct acagagacag atctggcagt ggaccaaccc aagatcttaa tacaatactc 2700
ctcacctcag cccaattatc ctctgatgtt gcagtggaga ctcccaaaca ggagtttgat 2760
gtatttggcg ctgtcccctt ctttgcagtg cgtgctcaac agccccagca agaaaagaat 2820
gaaaagaacc tccctcaaca caggtttcct gctgcaggac tggagcagga ggaatttgat 2880
gtattcacaa aggcgccttt tagcaagaag gtgaatgtac aagaatgcca tgcagtgggg 2940
cctgaggcac atactatccc tggttatccc aaaagtgtag atgtatttgg ctccactcca 3000
tttcagccct tcctcacatc aacaagtaaa agtgaaagca atgaggacct ttttgggctt 3060
gtgccctttg atgaaataac ggggagccag cagcaaaaaa gtcaaacagc gcagcttaca 3120
gaaactgtcc tctcgccaaa ggcgcacaaa gcaggatatg tccaaaagta atgggaagcg 3180
gcatcatggc acgccaacta gcacaaagaa gactttgaag cctacctatc gcactccaga 3240
gagggctcgc aggcacaaaa aagtgggccg cgagactctc aaagtagcaa tgaattttta 3300
accatctcag actccaagga gaacattagt gttgcactga ctgatgggaa agataggggg 3360
aatgtcttac aacctgagga gagcctgttg gacccccttcg gtgccaagcc cttccattct 3420
ccagacctgt catggcaccc tccacatcag ggcctgagcg acatccgtgc tgatcacaat 3480
actgtcctgc cagggcggcc aagacaaaat tcactacatg ggtcattcca tagtgcagat 3540
gtattgaaaa tggatgattt tggtgccgtg ccctttacag aacttgtggt gcaaagcatc 3600
actccacatc agtcccaaca gtcccaacca gtcgaattag acccatttgg tgctgctcca 3660
tttccttcta aacagtagat acttctgatg gattctcggc attaactcct gtttcaaaaa 3720
agtgtgaaca gttttatgaa tttgaaagaa aatttggtag ctctttatag cattcattct 3780
taaagatcag tcagaatagg tgatttctaa ataaaccaaa tagaagaatg aagtatctct 3840
acagggtagt aacttgattc ctcttcagga gaaaagggag ctaaattgca agctctaact 3900
aagggtttct gctactgaca tcacaacaca gaaatgcaag tgtggtactt ccagtgaaag 3960
cacatggcac ctttctaggt gtgtagccac tgagaaggga cagtgaaact gttatttttg 4020
atatcagaat gtcattttta tgtgcatatc cctaaaatta gggttatttc tacatacact 4080
agttacactt gtgaattttt tttaaggtct cttttaattt ccagacagtt aaaaacaatc 4140
tagttatctt aaagcattag aaagttatta tctggagagt gcagagattt cagtccatac 4200
acctttctcc acaaagcaga gccagaagta actgactatt gtgcctaaaa ctctgtttca 4260
tttttaaaaa caagtgccat taaaatggaa tatctaatga taagcatatg aaataatgtg 4320
taattagctc aatttaacta ttccacaact tacatattcc aaaacaatgt tatacatgat 4380
aaatatatat aatttttgtc agttaaaaca aattaaaaaa atggactatc gtcgcacaga 4440
agcctagaac aaaaatatga agagaaatat ctgacatttg taaagaaatt ataagaagaa 4500
aaaaagatac agaacagaaa acattcacta ctttagaaac actttatgca tggcttcttg 4560
ccccaaactt ttattgtgat ggccctaata aagcagatta ttggaaaaat tggaggacaa 4620
gggttgtata aaaattttat tttatgaaga aaatatgtag cggaaactga attttcaaga 4680
catttacaat gtgaaatcat gttgcattta acaatgtact ttattagcaa cttcaccaaa 4740
tattccccaa gtcataagca acaattattt ttattaggtt ttggggggtg gagtagtttt 4800
aataaagtgc acagaatggt gacacccaca aagccttata taaaggcagg attcatgcat 4860
cctgctgcaa gtacctctgc actaatatac cagatcctaa aatgcatata aggtggacta 4920
gcatcttaat tctgctagtt gattgtgtct ttactgaaaa gaacccagct accaatttgc 4980
cttttttttac accacaaatc ctaattagaa acttgaggtt ttatagaaat catttaatga 5040
tagagattac atatgtgaat taatgtgaat atagtatctg tgcttcctgt gtctatgact 5100
attttaagat ataattgtgc tgcgctatca gattaacatt tggaagtttc tagaacagtt 5160
aatgctattt acagaaagga gtagaaactc atcaactggc actctctttg atttttatat 5220
tttaaattaa ctctcttcga tctcaaagta tattttacga gtaattttat taggaatctc 5280
ttatagtgcc ccaatgggat aagctatttg cctattttca cagttctgaa cttggaaaga 5340
agcaaagtat atgtaactaa accacatatt tgtcttttta ttgcttttttc ccttctttta 5400
tatgctaaat caaatataga tttgtggata gggaagcaat atgtgaatca caatgtagca 5460
gaggcagacc aagcattaca ttattattta gagctggact gcaccaatta cttgtcctcg 5520
tgccaaaggc aaatatgttt gcaccttttt ttttttttct gattctcagg ttgattaata 5580
ctgctaatgc aaatgctcaa gtagatgttt aaaaacttta caaaatagat tcaagtgata 5640
ctttctttta aaagtgaaga gttgatgatt acacatagta aattcatgaa ctacagtagg 5700
tttgtatcaa acaatttttt ttaatgaaaa tctgttgagg tgtacacaat atgcttcttg 5760
attgtattag tccttggtct ctgctagacc tcatgagttt catcatttag aaaaggggta 5820
gaggatgaac taatgtctcc ttcagatgta aacatgaaat acctagagtt ttacttgctt 5880
ttcaatacac tgaataattt taaatgattc tgacactgat gtagaccctt tgacttataa 5940
attctgagga aacaactgac agcataaaat atttacattc ttataacaca gcacagtgac 6000
```

tttcttcttt caagattgta gctcagagaa aagatacagg attcaattgg gggttcaata 6060
ggatagaaat ggagagattc ctttgtgttg tagtagaggc attttcctaa ggagtataga 6120
tttatacttt gcattttcat tcatcatccc ccagaatcat ggtcaaggtg taggtcactc 6180
cacacagctg atgctcaggt tattcccttg tgagaattat gagaataaag ctcccaagat 6240
atgtgaaagt gcttaacaca gtacctggca cacagcactc aataaaagtt tggctctatt 6300
atgggatggt tcaattctgg tttaaggaag gaagaaaggt tattatatat gtaccactaa 6360
gcaaatatat atatatatat atatttgggg ttttttttcc ctaatattat ttgggtgtcc 6420
cctgtgcttc tttaggatgt agttataact aaacctgtta tacttgaaca tcactaagag 6480
aagtaaatta ttatgaagct agcaaaaatc ttgaggccaa agttgtttct taacagcttt 6540
aataatgctt gttgattttg aataatcctt taaaaagtgg accatttgct tattttaata 6600
tcacgtcagt aaaatgttag tattaaaaag atcagctttt tatggcattg aagaatgtat 6660
ctgctaagac acaaaaattg catggtaagt ataataggtg gaggaggaaa ggttgtaggc 6720
cggatgaaaa tttaactgac tagaacattt attcaggagt gtaattattt tcccttaccc 6780
caatccctgt gtacgtgttg ggtatagtta cgacattatc cggatttgca aatagacaca 6840
actttcagtc ttaccctgtt tattgtttaa gagtgataga ctgttgtcct cttgctggtg 6900
gaaaatctaa ggtggagccc actcttctat gctgaagttc accaggcaga gcagttttct 6960
tacaagtcag ctactctgct tggtttattt taggttttgg tacttcacgt aagcactgtt 7020
agaaggtaca agtgtattaa tatcactagt tttgaggcgc ttgggtacat ttgtttttaa 7080
tatatttaga atgtgcagta aacttttttc tcattttttt ttctttttag caaacttgtt 7140
attttaggtc caattattga gttgacagtc tactgtgaga atgagatgac atatctactg 7200
tgagaatacc ataaatgatg aatagtttat ttgagaactt ttatactcag tggtgtttat 7260
atattaagat aaaaatatgt acacacatgc atgtcacatc tctctactgt ggagttaatg 7320
tgaattttta aaaaatggaa ttgcaaccac aatcatatct aagagaacat tcactcctag 7380
tgagggttta caaaagctac taagagaata aggtagatga taatgcaaag ggtcatgatt 7440
tggggttatt tttgttttat tttaaaattt atactgctac ttttgaaaga attgttttta 7500
tgactatgct ctttttgtga ttgaaaagtc atctaataga agctgtatag aagctacttt 7560
ttaattgctg gcaaacagct ttaagtgcac tttctttgat tacacttcca ttttttgtta 7620
aacttgaatt ttctgaagcc ttttatgtac cactaagcaa ataactttaa cctttaaata 7680
aagcaaattt acatctttat tgtatttcta cttgttacaa aacatacttg ctaaagtaac 7740
ttcagtcctc aaatatagct gggaaacaat tatgagatag actcagtctc cccctcccac 7800
cccttttccc ctgccatatc taattaagca ctaactgatt ttattacttt attgccttta 7860
cactgcttat tctttttgac tgaattctgt ccctgattca ctgttttgtt tgaaatttaa 7920
agttattttc ttactgtatt tatcatacct gttttaatct gttttcttta aatgcaataa 7980
atcccaaatg gattgcatat tctttataat cagtg 8015

<210> 24
<211> 7945
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526214CB1

<400> 24
agccgggcag ctgcagcgga gccgcggagc gggcggcggg gcccaggctg tgcgcttggg 60
gagcgcggaa tgtgaggctt ggcgggccgc agcacgctcg gacgggccag gggcggcgac 120
ccctcgcgga cgcccggctg cgcgccgggc cggggacttg cccttgcacg ctccctgcgc 180
cctccagctc gccggcggga ccatgaagaa gttctctcgg atgcccaagt cggagggcgg 240
cagcggcggc ggagcggcgg gtggcggggc tggcggggcc ggggccgggg ccggctgcgg 300
ctccggcggc tcgtccgtgg gggtccgggt gttcgcggtc ggccgccacc aggtcaccct 360
ggaagagtcg ctggccgaag gtacgggcgc ccggggaggc tcggacaggc aggtggattc 420
tccacagttt tcctcgtgcg tactcacggt ggaatccgat gtgcattgaa gcgaatgtat 480
gtcaataaca tgccagacct caatgtttgt aaaagggaaa ttacaattat gaaagagcta 540
tctggtcaca aaaatattgt gggctatttg gactgtgctg ttaattcaat tagtgataat 600
gtatgggaag tccttatctt aatggaatat tgtcgagctg gacaggtagt gaatcaaatg 660
aataagaagc tacagacggg ttttacagaa ccagaagtgt tacagatatt ctgtgatacc 720
tgtgaagctg ttgcaaggtt gcatcagtgt aagactccaa taattcaccg ggatctgaag 780
gtagaaaata tttgctgaat gatggtggga actatgtact ttgtgacttt ggcagtgcca 840

ctaataaatt tcttaatcct caaaaagatg gagttaatgt agtagaagaa gaaattaaaa 900
agtatacaac tctgtcatac agagcccctg aaatgatcaa cctttatgga gggaaaccca 960
tcaccaccaa ggctgatatc tgggcactgg gatgtctact ctataaactt tgtttcttca 1020
ctcttccttt tggtgagagt caggttgcta tctgtgatgg caacttcacc atcccagaca 1080
attctcgtta ctcccgtaac atacattgct taataaggtt catgcttgaa ccagatccgg 1140
aacatagacc tgatatattt caagtgtcat attttgcatt taaatttgcc aaaaaggatt 1200
gtccagtctc caacatcaat aattcttcta ttccttcagc tcttcctgaa ccgatgactg 1260
ctagtgaagc agctgctagg aaaagccaaa taaaagccag aataacagat accattggac 1320
caacagaaac ctcaattgca ccaagacaaa gaccaaaggc caactctgct actactgcca 1380
ctcccagtgt gctgaccatt caaagttcag caacacctgt taaagtcctt gctcctggtg 1440
aattcggtaa ccatagacca aaggggcac taagacctgg aaatggccct gaaattttat 1500
tgggtcaggg acctcctcag cagccgccac agcagcatag agtactccag caactacagc 1560
agggagattg gagattacag caactccatt tacagcatcg tcatcctcac cagcagcagc 1620
agcagcagca gcagcaacag caacagcagc agcagcaaca gcaacagcag cagcagcagc 1680
agcagcagca gcagcagcac caccaccacc accaccaccc tacttcaaga tgcttatatg 1740
cagcagtatc aacatggcaa cacagcagca acagatgctt caacaacaat ttttaatgca 1800
ttcggtatat caaccacaac cttctgcatc acagtatcct acaatgatgc cgcagtatca 1860
gcaggctttc tttcaacagc agatgctagc tcaacatcag ccgtctcaac aacaggcatc 1920
acctgaatat cttacctccc ctcaagagtt ctcaccagcc ttagtttcct acacttcatc 1980
acttccagct caggttggaa ccataatgga ctcctcctat agtgccaata ggtcagttgc 2040
tgataaagag gccattgcaa atttcacaaa tcagaagaac atcagcaatc cacctgatat 2100
gtcagggtgg aatccttttg gagaggataa tttctctaag ttaacagaag aggaactatt 2160
ggacagagaa tttgaccttc taagatcaag ttctcctgaa aagaaagctg aacattcatc 2220
tataaatcaa gaaaatggca ctgcaaaccc tatcaagaac ggtaaaacaa gtccagcatc 2280
taaagatcag cggactggaa agaaaacctc agtacagggt caagtgcaaa aggggaatga 2340
tgaatctgaa agtgattttg aatcagatcc cccttctcct aagagcagtg aagaggaaga 2400
gcaagatgat gaagaagttc ttcaggggga acaaggagat tttaagtgat gatgatactg 2460
aaccagaaaa tctgggtcat aggcctctcc tcatggattc tgaagatgag gaagaagagg 2520
agaaacatag ctctgattct gattatgagc aggctaaagg caaagtacag tgacatgagc 2580
tctgtctaca gagacggatc tggcagtggg accaacccaa gatcttaata caatactcct 2640
cacctcagcc caattatcct ctgatgttgc agtggagact cccaaacagg agtttgatgt 2700
atttggcgct gtccccttct ttgcagtgcg tgctcaacag ccccagcaag aaaagaatga 2760
aaagaacctc cctcaacaca ggtttcctgc tgcaggactg gagcaggagg aatttgatgt 2820
attcacaaag gcgcctttta gcaagaaggt gaatgtacaa gaatgccatg cagtggggcc 2880
tgaggcacat actatccctg gttatcccaa aagtgtagat gtatttggct ccactccatt 2940
tcagcccttc ctcacatcaa caagtaaaag tgaaagcaat gaggaccttt ttgggcttgt 3000
gccctttgat gaaataacgg ggagccagca gcaaaaagtc aaacagcgca gcttacagaa 3060
actgtcctct cgccaaaggc gcacaaagca ggatatgtcc aaaagtaatg ggaagcggca 3120
tcatggcacg ccaactagca caaagaagac tttgaagcct acctatcgca ctccagagag 3180
ggctcgcagg cacaaaaaag tgggccgccg agactctcaa agtagcaatg aatttttaac 3240
catctcagac tccaaggaga acattagtgt tgcactgact gatgggaaag atagggggaa 3300
tgtcttacaa cctgaggaga gcctgttgga ccccttcggt gccaagccct tccattctcc 3360
agacctgtca tggcaccctc cacatcaggg cctgagcgac atccgtgctg atcacaatac 3420
tgtcctgcca gggcggccaa gacaaaattc actacatggg tcattccata gtgcagatgt 3480
attgaaaatg gatgattttg gtgccgtgcc ctttacagaa cttgtggtgc aaagcatcac 3540
tccacatcag tcccaacagt cccaaccagt cgaattagac ccatttggtg ctgctccatt 3600
tccttctaaa cagtagatac ttctgatgga ttctcggcat taactcctgt ttcaaaaaag 3660
tgtgaacagt tttatgaatt tgaaagaaaa tttggtagct ctttatagca ttcattctta 3720
aagatcagtc agaataggtg atttctaaat aaaccaaata gaagaatgaa gtatctctac 3780
agggtagtaa ctcgattcct cttcaggaga aaagggagct aaattgcaag ctctaactaa 3840
gggtttctgc tactgacatc acaacacaga aatgcaagtg tggtacttcc agtgaaagca 3900
catggcacct ttctaggtgt gtagccactg agaagggaca gtgaaactgt tatttttgat 3960
atcagaatgt catttttatg tgcatatccc taaaattagg gttatttcta catacactag 4020
ttacacttgt gaattttttt taaggtctct tttaatttcc agacagttaa aaacaatcta 4080
gttatcttaa agcattagaa agttattatc tggagagtgc agagatttca gtccatacac 4140
ctttctccac aaagcagagc cagaagtaac tgactattgt gcctaaaact ctgtttcatt 4200
tttaaaaaca agtgccatta aaatggaata tctaatgata agcatatgaa ataatgtgta 4260
attagctcaa tttaactatt ccacaactta catattccaa aacaatgtta tacatgataa 4320
atatatataa tttttgtcag ttaaaacaaa ttaaaaaaat ggactatcgt cgcacagaag 4380

cctagaacaa aaatatgaag agaaatatct gacatttgta aagaaattat aagaagaaaa 4440
aaagatacag aacagaaaac attcactact ttagaaacac tttatgcatg gcttcttgcc 4500
ccaaactttt attgtgatgg ccctaataaa gcagattatt ggaaaaattg gaggacaagg 4560
gttgtataaa aattttattt tatgaagaaa atatgtagcg gaaactgaat tttcaagaca 4620
tttacaatgt gaaatcatgt tgcatttaac aatgtacttt attagcaact tcaccaaata 4680
ttccccaagt cataagcaac aattattttt attaggtttt gggggtgga gtagttttaa 4740
taaagtgcac agaatggtga cacccacaaa gccttatata aaggcaggat tcatgcatcc 4800
tgctgcaagt acctctgcac taatatacca gatcctaaaa tgcatataag gtggactagc 4860
atcttaattc tgctagttga ttgtgtcttt actgaaaaga acccagctac caatttgcct 4920
tttttacac cacaaatcct aattagaaac ttgaggtttt atagaaatca tttaatgata 4980
gagattacat atgtgaatta atgtgaatat agtatctgtg cttcctgtgt ctatgactat 5040
tttaagatat aattgtgctg cgctatcaga ttaacatttg gaagtttcta gaacagttaa 5100
tgctatttac agaaaggagt agaaactcat caactggcac tctctttgat ttttatattt 5160
taaattaact ctcttcgatc tcaaagtata ttttacgagt aattttatta ggaatctctt 5220
atagtgcccc aatgggataa gctattgcct attttcacag ttctgaactt ggaaagaagc 5280
aaagtatatg taactaaagc acatattgtc ttttattgc tttttccctt cttttatatg 5340
ctaaatcaat atagatttgt ggatagggaa gcaatatgtg aatcacaatg tagcagaggc 5400
agaccaagca ttacattatt atttagagct ggactgcacc aattacttgt cctcgtgcca 5460
aaggcaaata tgtttgcacc tttttttttt tttctgattc tcaggttgat taatactgct 5520
aatgcaaatg ctcaagtaga tgtttaaaaa ctttacaaaa tagattcaag tgatactttc 5580
ttttaaaagt gaagagttga tgattacaca tagtaaattc atgaactaca gtaggtttgt 5640
atcaaacaat ttttttaat gaaaatctgt tgaggtgtac acaatatgct tcttgattgt 5700
attagtcctt ggtctctgct agacctcatg agtttcatca tttagaaaag gggtagagga 5760
tgaactaatg tctccttcag atgtaaacat gaaataccta gagttttact tgctttcaa 5820
tacactgaat aattttaaat gattctgaca ctgatgtaga ccctttgact tataaattct 5880
gaggaaacaa ctgacagcat aaaatattta cattcttata acacagcaca gtgactttct 5940
tctttcaaga ttgtagctca gagaaaagat acaggattca attgggggtt caataggata 6000
gaaatggaga gattcctttg tgttgtagta gaggcatttt cctaaggagt atagatttat 6060
actttgcatt ttcattcatc atcccccaga atcatggtca aggtgtaggt cactccacac 6120
agctgatgct caggttattc ccttgtgaga attatgagaa taaagctccc aagatatgtg 6180
aaagtgctta acacagtacc tggcacacag cactcaataa aagtttggct ctattatggg 6240
atggttcaat tctggtttaa ggaaggaaga aaggttatta tatatgtacc actaagcaaa 6300
tatatatata tatatatatt tggggttttt tttccctaat attatttggg tgtcccctgt 6360
gcttctttag gatgtagtta taactaaacc tgttatactt gaacatcact aagagaagta 6420
aattattatg aagctagcaa aaatcttgag gccaaagttg tttcttaaca gctttaataa 6480
tgcttgttga ttttgaataa tcctttaaaa agtggaccat ttgcttattt taatatcacg 6540
tcagtaaaat gttagtatta aaaagatcag ctttttatgg cattgaagaa tgtatctgct 6600
aagacacaaa aattgcatgg taagtataat aggtggagga ggaaaggttg taggccggat 6660
gaaaatttaa ctgactagaa catttattca ggagtgtaat tattttccct taccccaatc 6720
cctgtgtacg tgttgggtat agttacgaca ttatccggat ttgcaaatag acacaacttt 6780
cagtcttacc ctgtttattg tttaagagtg atagactgtt gtcctcttgc tggtggaaaa 6840
tctaaggtgg agcccactct tctatgctga agttcaccag gcagagcagt tttcttacaa 6900
gtcagctact ctgcttggtt tattttaggt tttggtactt cacgtaagca ctgttagaag 6960
gtacaagtgt attaatatca ctagttttga ggcgcttggg tacatttgtt tttaatatat 7020
ttagaatgtg cagtaaactt ttttctcatt tttttttctt tttagcaaac ttgttatttt 7080
aggtccaatt attgagttga cagtctactg tgagaatgag atgacatatc tactgtgaga 7140
ataccataaa tgatgaatag tttatttgag aacttttata ctcagtggtg ttttatatat 7200
taagataaaa atatgtacac acatgcatgt cacatctctc tactgtggag ttaatgtgaa 7260
tttttaaaaa atggaattgc aaccacaatc atatctaaga gaacattcac tcctagtgag 7320
ggtttacaaa agctactaag agaataaggt agatgataat gcaaagggtc atgatttggg 7380
gttatttttg ttttatttta aaatttatac tgctactttt gaaagaattg tttttatgac 7440
tatgctcttt ttgtgattga aaagtcatct aatagaagct gtatagaagc tactttttaa 7500
ttgctggcaa acagctttaa gtgcactttc tttgattaca cttccatttt ttgttaaact 7560
tgaattttct gaagcctttt atgtaccact aagcaaataa ctttaacctt taaataaagc 7620
aaatttacat ctttattgta tttctacttg ttacaaaaca tacttgctaa agtaacttca 7680
gtcctcaaat atagctggga aacaattatg agatagactc agtctccccc tcccacccct 7740
tttcccctgc catatctaat taagcactaa ctgattttat tactttattg cctttacact 7800
gcttattctt tttgactgaa ttctgtccct gattcactgt tttgtttgaa atttaaagtt 7860
attttcttac tgtatttatc atacctgttt taatctgttt tctttaaatg caataaatcc 7920

caaatggatt gcatattctt tataa 7945

<210> 25
<211> 3149
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526228CB1

<400> 25
ctcgcggtat catccggtgc tgaggccctg taataaaggt ctcgcgaaat ttgttctaga 60
ggtccaagtt gcttcttagc ttactccacc ccaccccc aa cctgtccctc cttttctttc 120
caagtcacaa aattctcccc tcccctaccc cggagtttac ggccctcctc ctgtttccga 180
tttcagcccg gaaccggaag tgaagtgggc ggggcccgtc ggcggaaaac gcagcggagc 240
cagagccgga cacggctgtg gccgctgcct ctacccccgc cacggatcgc cgggtagtag 300
gactgcgcgg ctccaggctg agggtcggtc cggaggcggg tgggcgcggg tctcacccgg 360
attgtccggg tggcaccgtt cccggcccca ccgggcgccg cgagggatca tgtctacagc 420
ctctgccgcc tcctcctcct cctcgtcttc ggccggtgag atgatcgaag ccccttccca 480
ggtcctcaac tttgaagaga tcgactacaa ggagatcgag gtggaagagg ttgttggaag 540
aggagccttt ggagttgttt gcaaagctaa gtggagagca aaagatgttg ctattaaaca 600
aatagaaagt gaatctgaga ggaaagcgtt tattgtagag cttcggcagt tatcccgtgt 660
gaaccatcct aatattgtaa agctttatgg agcctgcttg aatccagtgt gtcttgtgat 720
ggaatatgct gaagggggct ctttatataa tgtttgtgcc tttctttcgc agtgctgcat 780
ggtgctgaac cattgccata ttatactgct gcccacgcaa tgagttggtg tttacagtgt 840
tcccaaggag tggcttatct tcacagcatg caacccaaag cgctaattca cagggacctg 900
aaaccaccaa acttactgct ggttgcaggg gggacagttc taaaaatttg tgattttggt 960
acagcctgtg acattcagac acacatgacc aataacaagg ggagtgctgc ttggatggca 1020
cctgaagttt ttgaaggtag taattacagt gaaaaatgtg acgtcttcag ctggggtatt 1080
attctttggg aagtgataac gcgtcggaaa ccctttttgat gagattggtg gcccagcttt 1140
ccgaatcatg tgggctgttc ataatggtac tcgaccacca ctgataaaaa atttacctaa 1200
gcccattgag agcctgatga ctcgttgttg gtctaaagat ccttcccagc gcccttcaat 1260
ggaggaaatt gtgaaaataa tgactcactt gatgcggaga aatatttgct gttttcttca 1320
agactatcta gactccaatg taagctagat tcttagtact ttccaggagc agatgagcca 1380
ttacagtatc cttgtcagta ttcagatgaa ggacagagca actctgccac cagtacaggc 1440
tcattcatgg acattgcttc tacaaatacg agtaacaaaa gtgacactaa tatggagcaa 1500
gttcctgcca caaatgatac tattaagcgc ttagaatcaa aattgttgaa aaatcaggca 1560
aagcaacaga gtgaatctgg acgtttaagc ttggggagcc tcccgtggga gcagtgtgga 1620
gagcttgccc ccaacctctg agggcaagag gatgagtgct gacatgtctg aaatagaagc 1680
taggatcgcc gcaaccacag gcaacggaca gccaagacgt agatccatcc aagacttgac 1740
tgtaactgga acagaacctg gtcaggtgag cagtaggtca tccagtccca gtgtcagaat 1800
gattactacc tcaggaccaa cctcagaaaa gccaactcga agtcatccat ggacccctga 1860
tgattccaca gataccaatg gatcagataa ctccatccca atggcttatc ttacactgga 1920
tcaccaacta cagcctctag caccgtgccc aaactccaaa gaatctatgg cagtgtttga 1980
acagcattgt aaaatggcac aagaatatat gaaagttcaa acagaaattg cattgttatt 2040
acagagaaag caagaactag ttgcagaact ggaccaggat gaaaaggacc agcaaaatac 2100
atctcgcctg gtacaggaac ataaaaagct tttagatgaa aacaaaagcc tttctactta 2160
ctaccagcaa tgcaaaaaac aactagaggt catcagaagt cagcagcaga aacgacaagg 2220
cacttcatga ttctctggga ccgttacatt ttgaaatatg caaagaaaga ctttttttta 2280
aggaaaggaa aaccttataa tgacgattca tgagtgttag ctttttggcg tgttctgaat 2340
gccaactgcc tatatttgct gcattttttt cattgtttat tttccttttc tcatggtgga 2400
catacaattt tactgtttca ttgcataaca tggtagcatc tgtgacttga atgagcagca 2460
ctttgcaact tcaaaacaga tgcagtgaac tgtggctgta tatgcatgct cattgtgtga 2520
aggctagcct aacagaacag gaggtatcaa actagctgct atgtgcaaac agcgtccatt 2580
ttttcatatt agaggtggaa cctcaagaat gactttattc ttgtatctca tctcaaaata 2640
ttaataattt ttttcccaaa agatggtata taccaagtta aagacagggt attataaatt 2700
tagagtgatt ggtggtatat tacggaaata cggaaccttt agggatagtt ccgtgtaagg 2760
gctttgatgc cagcatcctt ggatcagtac tgaactcagt tccatccgta aaatatgtaa 2820

agataagcaa gatctaagaa gttatcaaaa ctattcttta aaatgctaaa gcagctcctg 2880
tagccagaga tcacaggtct tccctgtgaa actttggttt ctttctataa atgtgtgtgg 2940
ttttcagcgc tcaactcctg tcttcaaatg gtagtaagtt ctacttctac ttctgtcatt 3000
cagaacattt tatgtcaaat gatgtaatgc agaaattctt gtgcatattt gtaactgaag 3060
gaagcttttt agatttattt ttgtttttaa taaaattcag attcctattc taaactggta 3120
cataaaagtg gtgaatgact tgtatcagc 3149

<210> 26
<211> 3617
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526246CB1

<400> 26
taagatggcg gacctggagg cggtgctggc cgacgtgagc tacctgatgg ccatggagaa 60
gagcaaggcc acgccggccg cgcgcgccag caagaagata ctgctgcccg agcccagcat 120
ccgcagtgtc atgcagaagt acctggagga ccggggcgag gtgacctttg agaagatctt 180
ttcccagaag ctggggtacc tgctcttccg agacttctgc ctgaaccacc tggaggaggc 240
caggcccttg gtggaattct atgaggagat caagaagtac gagaagctgg agacggagga 300
ggagcgtgtg gcccgcagcc gggagatctt cgactcatac atcatgaagg agctgctggc 360
ctgctcgcat cccttctcga agagtgccac tgagcatgtc caaggccacc tggggaagaa 420
gcaggtgcct ccggatctct tccagccata catcgaagag atttgtcaaa acctccgagg 480
ggacgtgttc cagaaattca ttgagagcga taagttcaca cggttttgcc agtggaagaa 540
tgtggagctc aacatccacg tgagtgggct tggtggggc atggaaagcc acgcaccctg 600
ctgctcctct cccgggagct gggcctgtgg cttggctggg agggggaggt caggggatgt 660
ctgtccttta gcccccaggg ccgtggctat ggggggtcagg gccgggatcc cagcatgggg 720
aggccggagc aggtaaatat gtggcaagga tggccaggac atgggtatgg ggaccctggc 780
atggggccag cccctgctgc ccaggtgcct ctgccccagg gctgggcaga ggcagcctgt 840
ggtgaccgca gctgtcgctg cccctcagct gaccatgaat gacttcagcg tgcatcgcat 900
cattgggcgc gggggctttg gcgaggtcta tgggtgccgg aaggctgaca caggcaagat 960
gtacgccatg aagtgcctgg acaaaaagcg catcaagatg aagcaggggg agaccctggc 1020
cctgaacgag cgcatcatgc tctcgctcgt cagcactggg gactgcccat tcattgtctg 1080
catgtcatac gcgttccaca cgccagacaa gctcagcttc atcctggacc tcatgaacgg 1140
tggggacctg cactaccacc tctcccagca cggggtcttc tcagaggctg acatgcgctt 1200
ctatgcggcc gagatcatcc tgggcctgga gcacatgcac aaccgcttcg tggtctaccg 1260
ggacctgaag ggcacccac gggtacatgg ctccggaggt cctgcagaag ggcgtggcct 1320
acgacagcag tgccgactgg ttctctctgg ggtgcatgct cttcaagttg ctgcgggggc 1380
acagcccctt ccggcagcac aagaccaaag acaagcatga gatcgaccgc atgacgctga 1440
cgatggccgt ggagctgccc gactccttct cccctgaact acgctccctg ctggaggggt 1500
tgctgcagag ggatgtcaac cggagattgg gctgcctggg ccgagggggct caggaggtga 1560
aagagagccc ctttttccgc tccctggact ggcagatggt cttcttgcag aagtaccctc 1620
ccccgctgat ccccccacga ggggaggtga acgcggccga cgccttcgac attggctcct 1680
tcgatgagga ggacacaaaa ggaatcaagt tactggacag tgatcaggag ctctaccgca 1740
acttccccct caccatctcg gagcggtggc agcaggaggt ggcagagact gtcttcgaca 1800
ccatcaacgc tgagacagac cggctggagg ctcgcaagaa agccaagaac aagcagctgg 1860
gccatgagga agactacgcc ctgggcaagg actgcatcat gcatggctac atgtccaaga 1920
tgggcaaccc cttcctgacc cagtggcagc ggcggtactt ctacctgttc cccaaccgcc 1980
tcgagtggcg gggcgagggc gaggccccgc agagcctgct gaccatggag gagatccagt 2040
cggtggagga gacgcagatc aaggagcgca agtgcctgct cctcaagatc cgcggtggga 2100
aacagttcat tttgcagtgc gatagcgacc ctgagctggt gcagtggaag aaggagctgc 2160
gcgacgccta ccgcgaggcc cagcagctgg tgcagcgggt gcccaagatg aagaacaagc 2220
cgcgctcgcc cgtggtggag ctgagcaagg tgccgctggt ccagcgcggc agtgccaacg 2280
gcctctgacc cgcccacccg ccttttataa acctctaatt tattttgtcg aatttttatt 2340
atttgttttc ccgccaagcg gaaaaggttt tattttgtaa ttattgtgat ttcccgtggc 2400
cccagcctgg cccagctccc ccgggagggg cccgcttgcc tcggctcctg ctgcaccaac 2460
ccagccgctg cccggcgccc tctgtcctga cttcaggggc tgcccgctcc cagtgtcttc 2520

ctgtggggga agagcacagc cctcccgccc cttccccgag ggatgatgcc acaccaagct 2580
gtgccaccct gggctctgtg ggctgcactc tgtgcccatg ggcactgctg ggtggcccat 2640
ccccccctcac cagggcaggc acagcacagg gatccgactt gaattttccc actgcacccc 2700
ctcctgctgc agaggggcag gccctgcact gtcctgctcc acagtgttgg cgagaggagg 2760
ggcccgttgt ctccctggcc ctcaaggcct cccacagtga ctcgggctcc tgtgccctta 2820
ttcaggaaaa gcctctgtgt cactggctgc ctccactccc acttccctga cactgcgggg 2880
cttggctgag agagtggcat tggcagcagg tgctgctacc ctccctgctg tcccctcttg 2940
ccccaacccc cagcaccogg gctcagggac cacagcaagg cacctgcagg ttgggccata 3000
ctggcctcgc ctggcctgag gtctcgctga tgctgggctg ggtgccgccg cctcgcccac 3060
cgcatgcccc ctcgtgccag tcgcgctgcc tgtgtggtgt cgcgccttct ccccccccggg 3120
gctgggttgg cgcaccctcc cctcccgtct actcattccc cggggcgttt ctttgccgat 3180
ttttgaatgt gattttaaag agtgaaaaat gagactatgc gtttttataa aaaatggtgc 3240
ctgattcggc tgtctcagac tctttttgta cctggtgacc cctttcagc ttctgctggg 3300
ctggggcctg atggggaggg tctcggtggt accaggtctc ctccaccgcc atggcttcca 3360
aggtggtctg ctcgggccca ggccatcttc caggtggggt gaggcagtgg gtcccacttc 3420
ccctcctacc cctcccagct gacagtcctc tccacctagt ggctgtccag tgcccattcc 3480
tcacctttc ccggggagga gagagcagct tctgccactt cccaggtaag caggaggagg 3540
tgccaacagt gttaggcctg gcacagtgtc tgggctgact gggaccgtct caggcccaca 3600
gaacacccct gcacagc 3617

<210> 27
<211> 1955
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526258CB1

<400> 27
agtgcctgcc gggagccacg tctccgaaga ccgatagctg cttcgggatt ggcgtccggg 60
cggctatcta ggggctgctg ggaagatggc ggactcggtg gctagccgat gaggaggccg 120
cggggggaac ccggcccccg ggccccgaga ccgactgagg gagcgacctg cgcagggccc 180
ggggagtcat ggtctccatc acccaactcc atgcttcgag tcctgctctc tgctcagacc 240
tcccctgctc ggctgtctgg cctgctgctg atccctccag tacagccctg ctgtttgggg 300
cccagcaaat gggggaccg gcctgttgga ggaggcccca gtgcaggtcc tgtgcaagga 360
ctgcagcggc ttctggaaca ggcgaagagc cctggggagc tgctgcgctg gctgggccag 420
aaccccagca aggtgcgcgc ccaccactac tcggtggcgc ttcgtcgtct gggccagctc 480
ttggggtctc ggccacggcc ccctcctgtg gagcaggtca cactgcagga cttgagtcag 540
ctcatcatcc gtaactgccc ctcctttgac attcacacca tccacgtgtg tctgcacctt 600
gcagtcttac ttggctttcc atctgatggt cccctggtgt gtgccctgga acaggagcga 660
aggctccgcc tccctccgaa gccacctccc cctttgcagc cccttctccg agaggcaagg 720
ccagaggaac tgactcccca cgtgatggtg ctcctggccc agcacctggc ccggcaccgg 780
ttgcgggagc cccagcttct ggaagccatt acccacttcc tggtggttca ggaaacgcaa 840
ctcagcagca aggtggtaca gaagttggtc ctgccctttg ggcgactgaa ctacctgccc 900
ctggaacagc agtttatgcc ctgccttgag aggatcctgg ctcgggaagc aggggtggca 960
cccctggcta cagtcaacat cttgatgtca ctgtgccaac tgcggtgcct gcccttcaga 1020
gccctgcact ttgtttttc ccctggcttc atcaactaca tcagtggcac ccctcatgct 1080
ctgattgtgc gtcgctacct ctccctgctg gacacggccg tggagctgga gctcccagga 1140
taccggggtc cccgccttcc ccgaaggcag caagtgccca tctttcccca gcctctcatc 1200
accgaccgtg cccgctgcaa gtacagtcac aaggacatag tagctgaggg gttgcgccag 1260
ctgctggggg aggagaaata ccgccaggac ctgactgtgc ctccaggcta ctgcacaggt 1320
gagcaagggg caggcggcag gcccggggag acggagccct ggctaaggcc gccggccctg 1380
ctcccctcca gacttcctgc tgtgcgccag cagctctggt gctgtgcttc ccgtgaggac 1440
ccaggacccc ttcctgccat acccaccaag gtcctgccca cagggccagg ctgcctctag 1500
cgccactact cgagaccctg cccagaggta aaggaggcag ggtgggggag ccctggccac 1560
cttgcccgcc atgccctgag ccacgtccct ccccctgcag ggtggtgctg gtgttgcggg 1620
aacgctggca tttctgccgg gacggccggg tgctgctggg ctcgagggcc ctgagggagc 1680
ggcacctagg cctgatgggc taccagctcc tgccgctacc cttcgaggaa ctggagtccc 1740

agagaggcct gccccagctc aagagctacc tgaggcagaa gctccaggcc ctgggcctgc 1800
gctgggggcc tgaaggggc tgaggggttg atgtggggtt caggatggcc cccccatggg 1860
gggtggatga tttgcacttt ggttccctgt gttttgattt ctcattaaag ttcctttcct 1920
tccccgttgt gaatctcagt tttgggacgg ggagc 1955

<210> 28
<211> 2937
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526311CB1

<400> 28
gcgcaggggg ccgggctccg gctaggaggg tgggggccgc gccggtgaca gccgatcccc 60
gcccctgctg cccgccacgt ccctcacgta ccactcggca gaggcgcggg gaaacctggc 120
gtactggctg tggcttctct agcgggactc ggcatgaggc tggcgcggct gcttcgcgga 180
gccgccttgg ccggcccggg cccggggctg cgcgccgccg gcttcagccg cagcttcagc 240
tcggactcgg gctccagccc ggcgtccgag cgcggcgttc cgggccaggt ggacttctac 300
gcgcgcttct cgccgtcccc gctctccatg aagcagttcc tggacttcgg atcagtgaat 360
gcttgtgaaa agacctcatt tatgtttctg cggcaagagt tgcctgtcag actggcaaat 420
ataatgaaag aaataagtct ccttccagat aatcttctca ggacaccatc cgttcaattg 480
gtacaaagct ggtatatcca gagtcttcag gagcttcttg attttaagga caaaagtgct 540
gaggatgcta aagctattta tgaaaggcct agaagaacat ggttgcaggt ctctagttta 600
tgctgtatgg cctgcaagat gatctttatt gtttggtgga aaaggcaaag gaagtccatc 660
tcatcgaaaa cacattggaa gcataaatcc aaactgcaat gtacttgaag ttattaaaga 720
tggctatgaa aatgctaggc gtctgtgtga tttgtattat attaactctc ccgaactaga 780
acttgaagaa ctaaatgcaa aatcaccagg acagccaata caagtggttt atgtaccatc 840
ccatctctat cacatggtgt ttgaactttt caagaatgca atgagagcca ctatggaaca 900
ccatgccaac agaggtgttt accccctat tcaagttcat gtcacgctgg gtaatgagga 960
tttgactgtg aagatgagtg accgaggagg tggcgttcct ttgaggaaaa ttgacagact 1020
tttcaactac atgtattcaa ctgcaccaag acctcgtgtt gagacctccc gcgcagtgcc 1080
tctggctggt tttggttatg gattgcccat atcacgtctt tacgcacaat acttccaagg 1140
agacctgaag ctgtattccc tagagggtta cgggacagat gcagttatct acattaaggc 1200
tctgtcaaca gactcaatag aaagactccc agtgtataac aaagctgtct ggaagcatta 1260
caacaccaac cacgaggctg atgactggtg cgtcccccagc agagaaccca aagacatgac 1320
gacgttccgc agtgcctaga cacacttggg acatcggaaa atccaaatgt ggcttttgta 1380
ttaaatttgg aagtgtggcc cagagttgct cagaattgga gcagagcctg agacgtatct 1440
gcagatcctg tcatcagctg gcaagtccag gagactgtgt catttagaga ctgtgttgtt 1500
agttatccct caacatcttc taaggtggca ggaaataata ttggaaataa cattttaaag 1560
taaaaatttt aaagtttaaa gaagagtttt gccacttaaa caggggagct ttgtctggaa 1620
aatacactga gttgaaacac ttcatccttg gaaggattat ataagatgaa cagttgtgat 1680
aaatgtgtag attagaggga tgtgaatggg cagttagtcc agtgccctca tttaagaggc 1740
caagatcctg attcagagga ggcatccttt gcccagagct gcttagctaa tctgaccaaa 1800
tgttgggaaa aatgtctcac ctaacccact attccttaat tatggatttt gtgaaaaaca 1860
atagaacatg ttaatgagta atttatatta gttcgatgta ttacaatttt ttagctttaa 1920
attacagttt tcttataatg ttgaaatgtt ttagaatcct ttgaatctaa gtatttgttt 1980
cctaaatgaa acatttgtac aacatttgat gtttttactt atgaaatatt ctcctccccc 2040
aagaaaattt aaactttttc tctctattta aaagctaaga aatgttttaa aggaaaaatg 2100
aaattatctt cctttagctt atttttaagg taaaacagct ttttactctg ttattgtggt 2160
aatggacaga atattacata caaaaatatt ctgggagagc tttttcctag ttggttttaa 2220
atcattgtgc cacctgaaag gtttttagat tttataggag ctaatttgtc caccagcatt 2280
aatgtaacac agtgtagtta tgaaaatata ttgaaggaca ggaagtggac acgaagtgat 2340
ttttgtaacc tgagcagtta atgaatgtgc caacatttc taggaaggga cagcaagaat 2400
attctgctct gtagttaaaa tactggctgg cttttgatgt cttcatgctt aattgtgatc 2460
actttcttgc actgtgatgt ttttacgtga atatgttgaa gtagaagtct accatattat 2520
tttataaaat gttttctgta tggcaataaa ctgaaaacat ggatcaaccc ttcttttgaa 2580
aataaactga gtcaatttag ccttttaaaa atatagtcat ctcttttaaa tagaatcctc 2640

ttccaccatc aaggctcaac attttgtaag catccaaaaa attggtaatt agggggcttg 2700
cactaaattt cactatcttc agtagagagg aactgtttgg aacttagatt tccaatgtgt 2760
atattctaat ggagaaagca agaggtagag tttgtatgtt tgacttacct tagattttta 2820
ttttccatac atactgcaaa tgattgactt gttgcataaa tgaagatctt ctgttgtgtg 2880
cttttcaaac actgtaaata aatttgaaat ttgaataact ttccacagta taactgt 2937

<210> 29
<211> 6122
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526315CB1

<400> 29
ctgagaggag acctggtgac acagtagcct ttcggcagag ctccttggga tgagtaggaa 60
gtgctgctgc aggttttgtc tgggggatat ctgagccatt tctctgtggg cagctgtgtt 120
tcaaagtctg ggcaggttgt tgttgaattt tgcgtgggct gccaggattt tgtggaagta 180
taatactttg tcattatgag atgtcgtctc tcggtgcctc ctttgtgcaa attaaatttg 240
atgacttgca gtttttttgaa aactgcggtg gaggaagttt tgggagtgtt tatcgagcca 300
aatggatatc acaggacaag gaggtggctg taaagaagct cctcaaaata gagaaagagg 360
cagaaatact cagtgtcctc agtcacagaa acatcatcca gttttatgga gtaattcttg 420
aacctcccaa ctatggcatt gtcacagaat atgcttctct gggatcactc tatgattaca 480
ttaacagtaa cagaagtgag gagatggata tggatcacat tatgacctgg gccactgatg 540
tagccaaagg aatgcattat ttacatatgg aggctcctgt caaggtgatt cacagagacc 600
tcaagtcaag aaacgttgtt atagctgctg atggagtatt gaagatctgt gactttggtg 660
cctctcggct ccataaccat acaacacaca tgtccttggt tggaactttc ccatggatgg 720
ctccagaagt tatccagagt ctccctgtgt cagaaacttg tgacacatat tcctatggtg 780
tggttctctg ggagatgcta acaagggagg tcccctttaa aggtttggaa ggattacaag 840
tagcttggct tgtagtggaa aaaaacgaga ggctaaagaa actagagcgt gatctcagct 900
ttaaggagca ggagcttaaa gaacgagaaa gacgtttaaa gatgtgggag caaaagctga 960
cagagcagtc caacaccccg cttctcttgc ctcttgttgc aagaatgtct gaggagtctt 1020
actttgaatc taaaacagag gagtcaaaca gtgcagagat gtcatgtcag atcacagcaa 1080
caagtaacgg ggagggccat ggcatgaacc caagtctgca ggccatgatg ctgatgggct 1140
ttggggatat cttctcaatg aacaaagcag gagctgtgat gcattctggg atgcagataa 1200
acatgcaagc caagcagaat tcttccaaaa ccacatctaa gagaaggggg aagaaagtca 1260
acatggctct ggggttcagt gattttgact tgtcagaagg tgacgatgat gatgatgatg 1320
acggtgagga ggaggataat gacatggata atagtgaatg aaagcagaaa gcaaagtaat 1380
aaaatcacaa atgtttggaa aacacaaaag taacttgttt atctcagtct gtacaaaaac 1440
agtaaggagg cagaaagcca agcactgcat ttttaggcca atcacattta catgaccgta 1500
atttcttatc aattctactt ttatttttgc ttacagaaaa acggggggag aattaagcca 1560
aagaagtaca tttatgaatc agcaaatgtg gtgcctgatt atagaaattt gtgatcctat 1620
atacaatata ggacttttaa agttgtgaca ttctggcttt ttcttttaat gaatactttt 1680
tagtttgtat ttgactttat ttcctttatt caaatcattt ttaaaaactt acattttgaa 1740
caaacactct taactcctaa ttgttctttg acacgtagta attctgtgac atactttttt 1800
tttcttatag caatacactg taatatcaga aatggttggc ctgagcaacc tagtaagacc 1860
tcgtctctac taataattaa aaaactagct ggcatggtag cacacacctg tagtcccaga 1920
tacttgggag gccaaggcag gaggattgct tgagacctag caatcagtca gggctgcagt 1980
gagccatgat ggcaccactg cactctagcc tgggcaagag aacaagatcc tgtctcaaaa 2040
aacaaaaaaa agaaagaatt gatagtacaa aatccaacaa caatactgag atgatctaag 2100
aaggttataa caaaatgctc ttcagaaata cctaagtgct gagaattttt agtactaaag 2160
agcacagctg ctcaaagtaa agcctgagca gtgttctcag taatgtattt gaaggaaaaa 2220
taccctgatt tgaaaccaac agcagatgtt gcaaactttc ataccactgc tggccatgga 2280
agcctcttaa caacacactg tcatttaagg ctgtgcttgt gctttataca aagagaaaga 2340
ggtggtctta aggggatgct tccagggggt gagttcatgc ctctcctgta ttttccagca 2400
agtggggtat gtgtggtggt tgttttttag aggggcataa taatccagga ttctaagcat 2460
atgctcagct attttaaaga ggaaattaaa tattataaaa gaaatagtaa agataagtta 2520
tcctcactta ggcaaaagca caggtccttt ccatatcaag tttagcctac cagggttgtt 2580

ttttgtttta accctgctta ataatgttgg tgttttagaa gtagatacag gcactgctct 2640
gaaaacctgg ctagccaagg atattctcag aatgttatca cctgtttgtc aaagcttgtt 2700
taaattataa aacactttta attatatata tgaggcaaaa gaactaagac tttttcaaa 2760
ctaaattaga aaggagtgtc attatttgac tgttaaacca aaatattttt ggtgggtctt 2820
tttatggaag tttaaagaaa ggacatcatc atagatatga tctaacagta tttctaacta 2880
tatttgatca ttaaaagcct cttggaattt gaagcgtgac gtgtttctaa tgccccttga 2940
gaggtgaaaa ataccacata atgatcagta tgctgtgcca gcttcatttg gggagaaata 3000
actagtagaa agttctgggt gtgaggtgta cagcagtcta ggtggcatag tgatgaagaa 3060
agggatcaga gtctgactgt cactcagaat cctgggctca gttgcttgac aaccttggga 3120
aaattgtttt atctttgtgc gtctgtttgc tgatcttcag cgtgggaata ataacagtac 3180
ctacttgaaa ggatcattgt gcggattaaa agaaataata tatgtaaagc actttaacac 3240
agcaccaggc ccacggaaag tggctaatgt tagctactat gaatggtgcc agtgaagaca 3300
ctgaaaaata agtgatttca gtaaccttct ggaaagctat cagtttcaaa taatatttc 3360
tctgtagtat gagatgaaat taaaagtgga tagctttcag gaaagataaa gagaacatgc 3420
ttagaatgta agctaaacag atttttctg ttgctctttg aaaactatga gccctggcca 3480
gcttaacctg gtctgaggtg agactaaaca caaaaacagt agataaatct ctccctaaaa 3540
gatggattcc cccacatacc catgctacta gtttctctgt ctattcacac atatgtacaa 3600
atacatgaac acagcctgtc tgtgctcaga catagagaag tactacctga cttgagtcaa 3660
tgcacccaag aagaaaagct tggagtagag cagaagggag ggcttgggac tcctgtcttt 3720
ccagcatgcc ctggggtgca gtggtcagcc acctgaagag agagccaata gccatggggt 3780
ttacaaggca aagatagtca ttcattcaaa cacatattca tagaagctcc ttctctgtgc 3840
cagacaactg ttctggaaga tagctagatg aaaatctttg cactcacagg agcttaacat 3900
gccagtgagt gaagatcgat gataaataaa gcaaatgcat catatgttca catttgataa 3960
gtatatgcca aaaaatgaag ccgggaagga ggacaaggcc catgggtggg tgttgaggtt 4020
tttaaagtgt ggtcaggaaa ggccccactg ataaggtaac atttgagcaa gtctgaaaaa 4080
ggcaagggga tctttggggc taacttcggg atccctgcac tttatgtaag aatgtaaacc 4140
tggagtctca tttaagaatg atcagcaata cgtttagaac atatgaactg aatgaaatgg 4200
acattttttc ttaatttatg tataaatcca tatgattata cataaagttc tgatgcatta 4260
ataaaagcag ccaaataggg ccaaagagaa aaataacagg actctgtact ggacctaact 4320
ttatcattaa ttaggtaata ttttcctcat ttctttactg ctgccatttt cctcaccagt 4380
attccagaga tggtcatagc tcattactct accaccaaga acctaaaagg aattagaata 4440
cagcagaatt ggcctcagtg aagagcttaa aattgttctc ctcgtagaac tggactattg 4500
atcattacca cgtgacgttg gctctattac tttctgttcc caatgtcctt ctagtggttt 4560
gaaaatgtta aaacatccaa aaaaaaacaa cccggtagca ttgtcccttc cccactgaca 4620
aacttatcaa atccagaagc tttagagttt cgtctctaat tatttttctc ctgaacaaaa 4680
ttacccaagt caaaacaaaa tgtattttta gaattacggc agcatacgac ctgaattttg 4740
tgagtttcgt ggctttatct taaatcacca tttccctaaa aatggtttct ttctccttag 4800
aaatgctggt ggcaacttga tgaaacagcc aaatgcacca gggcaggtca ctttcccatt 4860
acactgattc cacaattaaa aaaataaaaa aaagaaaaaa aactcattga gatagctaca 4920
gttctatagg ttaatttaaa gcctcctttt tctactcatt tttgaaagca aaattacatt 4980
ttactatttt acataaccag tgaaaagacg ttgaaagcct acagctcact gttttgggtg 5040
ctctggaaat gttgagggtg ggtttttaac cagtgatttt taacgtgcag tgaatttgtt 5100
agacttttaa acaccagcta aggtagtcaa acttgatccc cattaaaaat caaggaatta 5160
ggggtcgggg gagggtttag gagtgatcca gaatgacctc ccagaattac tgtgcgtaca 5220
actttatttt tcagagtttt cattgggaat ggtaagagtg tttatgaaag acagttttaa 5280
aacttattct gagttaaata ttaatacttt aaaaaattat tgtactagac ttatcgcagc 5340
cttttgaaag tagcagagtt tcatcatacc acatatataa cagagcataa atttctata 5400
atcaggcacc ttttgctgct tttgagtaag actgttttcc tgtttaagtg ttaagcatcg 5460
ccagacataa aaatctattc tctcctctcg attgtagcat agcctgacag ctctagatac 5520
agcatttcta tgatgaaaaa tgagtatcca tcaggaaatc tagaagacta gccgtgtttt 5580
ctcagactcc acctttgttt gcactctgtt gcctgtgagg agctttctgg catgtgatta 5640
tttacttcaa aactagagtt ccaagcacct acattaatta ttttatattg tgtgcagaat 5700
agtatatctt ttaatgtcag atatgataca ctgcacatat tgcttttgca ctcttaaaat 5760
ttttgtacta aataatagaa aatatttata ttctttgagt gtgagctttg aatagatggc 5820
attatcactt tattgttttt ttaacaaaaa ctttttctca attattctat tgcaatgtta 5880
ttctgagcaa gtcctatgcc aaatatcttg tataatgttt gtatggaaga ttaaatttta 5940
ctcttgtgtg gtaagactat ttcagttact gattttatag ttggaatttg atattccagc 6000
acaaagtcca cagtgtattc agaaatccaa gttggtgtca tacatttcat tttgatgtga 6060
acttttcttt gctttccttt gttctaagac tccattttgc aataaacgtt ttgacagcaa 6120

aa 6122

<210> 30
<211> 1914
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7526442CB1

<400> 30
gctgtcatcg ttccgtgggc cctgctgcgg gcacgctctc ggcgcatgcg tttttatgc 60
gggattaagc ttgctgctgc gtgacagcgg agggctagga aaaggcgcag tggggcccgg 120
agctgtcacc cctgactcga cgcagcttcc gttctcctgg tgacgtcgcc tacaggaacc 180
gccccagtgg tcagctgccg cgctgttgct aggcaacagc gtgcgagctc agatcagcgt 240
ggggtggagg agaagtggag tttggaagtt caggggcaca ggggcacagg cccacgactg 300
cagcgggatg gaccagtact gcatcctggg ccgcatcggg gagggcgccc acggcatcgt 360
cttcaaggcc aagcacgtgg agactggcga gatagttgcc ctcaagaagg tggccctaag 420
gcggttggaa gacggcttcc ctaaccaggc cctgcgggag attaaggctc tgcaggagat 480
ggaggacaat cagtatgtgg tacaactgaa ggctgtgttc ccacacggtg gaggctttgt 540
gctggccttt gagttcatgc tgtcggatct ggccgaggtg gtgcgccatg cccagaggcc 600
actagcccag gcacaggtca agagctacct gcagatgctg ctcaagggtg tcgccttctg 660
ccatgccaac aacattgtac atcgggacct gcccccaagg cccatccagg gcccccccac 720
atccatgact tccacgtgga ccggcctctt gaggagtcgc tgttgaaccc agagctgatt 780
cggcccttca tcctggaggg gtgagaagtt ggccctggtc ccgtctgcct gctcctcagg 840
accactcagt ccacctgttc ctctgccacc tgcctggctt caccctccaa ggcctcccca 900
tggccacagt gggcccacac cacaccctgc cccttagccc ttgcgagggt tggtctcgag 960
gcagaggtca tgttcccagc caagagtatg agaacatcca gtcgagcaga ggagattcat 1020
ggcctgtgct cggtgagcct taccttctgt gtgctactga cgtacccatc aggacagtga 1080
gctctgctgc cagtcaaggc ctgcatatgc agaatgacga tgcctgcctt ggtgctgctt 1140
ccccgagtgc tgcctcctgg tcaaggagaa gtgcagagag taaggtgtcc ttatgttgga 1200
aactcaagtg gaaggaagat ttggtttggt tttattctca gagccattaa acactagttc 1260
agtatgtgag atatagattc taaaaacctc aggtggctct gccttatgtc tgttcctcct 1320
tcatttctct caagggaaat ggctaaggtg gcattgtctc atggctctcg ttttttggggt 1380
catggggagg gtagcaccag gcatagccac ttttgccctg agggactcct gtgtacttca 1440
catcactgag cactcattta gaagtgaggg agacagaagt ctaggcccag ggatggctcc 1500
agttggggat ccagcaggag accctctgca catgaggctg gtttaccaac atctactccc 1560
tcaggatgag cgtgagccag aagcagctgt gtatttaagg aaacaagcgt tcctggaatt 1620
aatttataaa tttaataaat cccaatataa tcccagctag tgctttttcc ttattataat 1680
ttgataaggt gattataaaa gatacatgga aggaagtgga accagatgca gaagaggaaa 1740
tgatggaagg acttatggta tcagatacca atatttaaaa gtttgtataa taataaagag 1800
tatgattgtg gttcaaggat aaaaacagac tagagaaact tattcttagc catcctttat 1860
ttttatttta tttatttttt gatggagtct tgcactccag cctggtgaca gact 1914

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
7526185	1	7526185CD1	16	7526185CB1
7526192	2	7526192CD1	17	7526192CB1
7526193	3	7526193CD1	18	7526193CB1
7526196	4	7526196CD1	19	7526196CB1
7526198	5	7526198CD1	20	7526198CB1
7526208	6	7526208CD1	21	7526208CB1
7526212	7	7526212CD1	22	7526212CB1
7526213	8	7526213CD1	23	7526213CB1
7526214	9	7526214CD1	24	7526214CB1
7526228	10	7526228CD1	25	7526228CB1
7526246	11	7526246CD1	26	7526246CB1
7526258	12	7526258CD1	27	7526258CB1
7526311	13	7526311CD1	28	7526311CB1
7526315	14	7526315CD1	29	7526315CB1
7526442	15	7526442CD1	30	7526442CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	7526185CD1	g2582413	8.0E-74	STE20-like kinase 3 [Homo sapiens]
				Schinkmann, K. A. et al., Cloning and characterization of a novel mammalian STE20-like kinase (mst-3), J. Biol. Chem. 272, 286995-286703 (1997)
1	7526185CD1	336486|STK24	4.4E-75	[Homo sapiens][Protein kinase;Transferase] Serine-threonine kinase 24 (Ste20 yeast homolog), member of the SPS1 subgroup of the STE20-like protein family, a serine-threonine kinase that prefers manganese as a cofactor and uses either GTP or ATP as a phosphate donor
				Zhou, T. H. et al., Identification of a human brain-specific isoform of mammalian STE20-like kinase 3 that is regulated by cAMP-dependent protein kinase., J Biol Chem 275, 2513-9 (2000).
1	7526185CD1	743574|MST4	4.0E-65	[Homo sapiens][Protein kinase;Transferase] Mst3 and SOK1-related kinase, a protein kinase, induces apoptosis, involved in cell growth, appears to activate MAPK but not JNK nor p38 kinase pathways, alternative form MST4a may regulate MST4; gene maps to a region associated with mental retardation
				Lin, J. L. et al., MST4, a new Ste20-related kinase that mediates cell growth and transformation via modulating ERK pathway. Oncogene 20, 6559-69. (2001).
2	7526192CD1	g2199529	1.5E-134	casein kinase I gamma 2 [Homo sapiens]
				Kitabayashi, A. N. et al., Cloning and chromosomal mapping of human casein kinase I gamma 2 (CSNK1G2), Genomics 46, 133-137 (1997)
2	7526192CD1	344104|CSNK1G2	8.1E-136	[Homo sapiens][Protein kinase;Transferase] Casein kinase 1 gamma 2, a putative serine/threonine protein kinase, may play a role in signal transduction
				Kitabayashi, A. N. et al., Cloning and chromosomal mapping of human casein kinase I gamma 2 (CSNK1G2)., Genomics 46, 133-7 (1997).
2	7526192CD1	664931|Csnk1g2	2.9E-129	[Rattus norvegicus][Protein kinase;Transferase] Casein kinase 1 gamma 2, serine/threonine protein kinase, may play a role in receptor tyrosine kinase-mediated signal transduction
				Voisin, L. et al., Angiotensin II stimulates serine phosphorylation of the adaptor protein Nck: physical association with the serine/threonine kinases Pak1 and casein kinase I., Biochem J 341, 217-23 (1999).
3	7526193CD1	g15215576	1.1E-166	BMP-2 inducible kinase [Mus musculus]

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Kearns, A. E. et al., Cloning and characterization of a novel protein kinase that impairs osteoblast differentiation in vitro, J. Biol. Chem. 276, 42213-42218 (2001)
3	7526193CD1	770160|Bike	6.1E-168	[Mus musculus] Protein containing a protein kinase domain, has low similarity to C. elegans SEL-5, which is a serine-threonine protein kinase that likely regulates LIN-12 and GLP-1 signaling
				Kearns, A. E. et al. (supra)
3	7526193CD1	244458|sel-5	1.2E-60	[Caenorhabditis elegans][Protein kinase][Cytoplasmic] Serine/threonine protein kinase which likely regulates LIN-12 and GLP-1 signaling; has similarity to S. cerevisiae Ark1p and Prk1p protein kinases which are involved in regulation of the cytoskeleton
				Fares, H. et al., SEL-5, A Serine/Threonine Kinase That Facilitates lin-12 Activity in Caenorhabditis elegans., Genetics 153, 1641-1654 (1999).
4	7526196CD1	g2506080	4.5E-40	HsGAK [Homo sapiens]
				Kimura, S. H. et al., Structure, expression, and chromosomal localization of human GAK, Genomics 44, 179-187 (1997)
4	7526196CD1	342050|GAK	2.5E-41	[Homo sapiens][Protein kinase;Transferase] Cyclin G-associated kinase, a putative serine/threonine protein kinase that shares homology with tensin and auxilin, may play a role in cell cycle regulation
				Kimura, S. H. et al. (supra)
4	7526196CD1	704892|Gak	1.1E-40	[Rattus norvegicus][Protein kinase;Transferase] Cyclin G-associated kinase, a serine/threonine protein kinase that shares homology with tensin and auxilin, interacts with cyclin G (Ccng1)- Cdk5 complex, involved in the dissociation of clathrin-coated vesicles in non-neuronal cells
				Greener, T. et al., Role of cyclin G-associated kinase in uncoating clathrin-coated vesicles from non-neuronal cells., J Biol Chem 275, 1365-70. (2000).
5	7526198CD1	g2506080	0.0	HsGAK [Homo sapiens]
				Kimura, S. H. et al. (supra)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
5	7526198CD1	342050|GAK	0.0	[Homo sapiens][Protein kinase;Transferase] Cyclin G-associated kinase, a putative serine/threonine protein kinase that shares homology with tensin and auxilin, may play a role in cell cycle regulation
				Kimura, S. H. et al. (supra)
5	7526198CD1	704892|Gak	0.0	[Rattus norvegicus][Protein kinase;Transferase] Cyclin G-associated kinase, a serine/threonine protein kinase that shares homology with tensin and auxilin, interacts with cyclin G (Ccng1)- Cdk5 complex, involved in the dissociation of clathrin-coated vesicles in non-neuronal cells
				Greener, T. et al. (supra)
6	7526208CD1	g4426595	9.0E-255	multifunctional calcium/calmodulin-dependent protein kinase II delta2 isoform [Homo sapiens]
				Hoch, B. et al., Identification and expression of delta-isoforms of the multifunctional Ca2+/calmodulin-dependent protein kinase in failing and nonfailing human myocardium, Circ. Res. 84, 713-721 (1999)
6	7526208CD1	742886|CAMK2D	4.9E-256	[Homo sapiens][Protein kinase;Transferase][Nuclear;Cytoplasmic] Calcium/calmodulin-dependent protein kinase II delta, member of the multifunctional CAMKII family involved in Ca2+ regulated processes; alternative form delta 3 is specifically upregulated in the myocardium of patients with heart failure
				Hoch, B. et al.(supra)
6	7526208CD1	772372|Camk2d	3.1E-243	[Mus musculus] Protein with strong similarity to calcium-calmodulin-dependent protein kinase II delta (rat Camk2d), which is involved in Ca2+ regulated processes, contains two protein kinase domains
				Hoch, B. et al., delta-Ca(2+)/calmodulin-dependent protein kinase II expression pattern in adult mouse heart and cardiogenic differentiation of embryonic stem cells, J Cell Biochem 79, 293-300 (2000).
7	7526212CD1	g1661132	5.3E-169	calcium/calmodulin-dependent protein kinase II delta 2-subunit [Sus scrofa]

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Singer, H. A. et al., Novel Ca2+/calmodulin-dependent protein kinase II gamma-subunit variants expressed in vascular smooth muscle, brain, and cardiomyocytes, J. Biol. Chem. 272, 9393-9400 (1997)
7	7526212CD1	772372|Camk2d	2.9E-170	[Mus musculus] Protein with strong similarity to calcium-calmodulin-dependent protein kinase II delta (rat Camk2d), which is involved in Ca2+ regulated processes, contains two protein kinase domains
				Hoch, B. et al., J Cell Biochem 79, 293-300 (2000). (supra)
7	7526212CD1	742886|CAMK2D	1.6E-169	[Homo sapiens][Protein kinase;Transferase][Nuclear;Cytoplasmic] Calcium/calmodulin-dependent protein kinase II delta, member of the multifunctional CAMKII family involved in Ca2+ regulated processes; alternative form delta 3 is specifically upregulated in the myocardium of patients with heart failure
				Hoch, B. et al., Circ Res 84, 713-21. (1999). (supra)
8	7526213CD1	g15215576	2.1E-15	BMP-2 inducible kinase [Mus musculus]
				Kearns, A. E. et al. (supra)
8	7526213CD1	605792|BIKE	1.7E-27	[Homo sapiens][Protein kinase;Transferase] Protein containing a eukaryotic protein kinase domain
8	7526213CD1	770160|Bike	1.1E-16	[Mus musculus] Protein containing a protein kinase domain, has low similarity to C. elegans SEL-5, which is a serine-threonine protein kinase that likely regulates LIN-12 and GLP-1 signaling
				Kearns, A. E. et al. (supra)
9	7526214CD1	g15215576	1.7E-16	BMP-2 inducible kinase [Mus musculus]
				Kearns, A. E. et al. (supra)
9	7526214CD1	605792|BIKE	3.8E-28	[Homo sapiens][Protein kinase;Transferase] Protein containing a eukaryotic protein kinase domain
9	7526214CD1	770160|Bike	9.4E-18	[Mus musculus] Protein containing a protein kinase domain, has low similarity to C. elegans SEL-5, which is a serine-threonine protein kinase that likely regulates LIN-12 and GLP-1 signaling
				Kearns, A. E. et al. (supra)
10	7526228CD1	g2924624	4.6E-55	TGF-beta activated kinase 1a [Homo sapiens]

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Sakurai, H. et al., TGF-beta-activated kinase 1 stimulates NF-kappa B activation by an NF-kappa B-inducing kinase-independent mechanism, Biochem. Biophys. Res. Commun. 243, 545-549 (1998)
10	7526228CD1	338400|MAP3K7	2.5E-56	[Homo sapiens][Protein kinase;Transferase] Mitogen-activated protein kinase kinase kinase 7 (TGF beta activated kinase 1), mediates TGFbeta and IL1 signal transduction, induces NFkappaB activation, may act as a regulatory kinase of I kappa B kinases (IKKs)
				Sakurai, H. et al., Functional interactions of transforming growth factor beta-activated kinase 1 with IkappaB kinases to stimulate NF-kappaB activation., J Biol Chem 274, 10641-8 (1999).
10	7526228CD1	338400|MAP3K7	2.50E-56	[Homo sapiens][Protein kinase;Transferase] Mitogen-activated protein kinase kinase kinase 7 (TGF beta activated kinase 1), mediates TGFbeta and IL1 signal transduction, induces NFkappaB activation, may act as a regulatory kinase of I kappa B kinases (IKKs)
				Craig, R. et al., p38 MAPK and NF-kappa B collaborate to induce interleukin-6 gene expression and release. Evidence for a cytoprotective autocrine signaling pathway in a cardiac myocyte model system., J Biol Chem 275, 23814-24 (2000).
11	7526246CD1	g23272739	5.7E-96	adrenergic, beta, receptor kinase 1 [Homo sapiens]
				Strausberg, R. L. et al., Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences, Proc. Natl. Acad. Sci. U.S.A. 99, 16899-16903 (2002)
11	7526246CD1	334086|ADRBK 1	3.1E-97	[Homo sapiens][Protein kinase;Transferase][Cytoplasmic;Plasma membrane] Beta-adrenergic receptor kinase 1, kinase that mediates desensitization of G protein-coupled receptors, phosphorylated by PKC, may modulate cardiovascular function; mouse and rat Adrbk1 appear to be involved with cardiomyopathy and myocardial infarction
				Shih, M. et al., Oligodeoxynucleotides antisense to mRNA encoding protein kinase A, protein kinase C, and beta-adrenergic receptor kinase reveal distinctive cell-type-specific roles in agonist-induced desensitization., Proc Natl Acad Sci U S A 91, 12193-7 (1994).
11	7526246CD1	775647|Adrbk1	1.1E-94	[Mus musculus][Protein kinase;Transferase] Beta-adrenergic receptor kinase 1, a kinase that may mediate desensitization of G protein-coupled receptors, modulates myocardial function and involved in cardiomyopathy; human ADRBK1 may play roles in hypertension and cardiomyopathy

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Proll, M. A. et al., Beta 2-adrenergic receptor mutants reveal structural requirements for the desensitization observed with long-term epinephrine treatment., Mol Pharmacol 44, 569-74 (1993).
12	7526258CD1	g33303889	9.6E-110	FAST kinase [synthetic construct]
12	7526258CD1	743544|FASTK	5.2E-111	[Homo sapiens][Protein kinase;Transferase] Fas-activated serine threonine kinase, a serine-threonine kinase that phosphorylates RNA binding protein TIA1 during Fas mediated apoptosis, upregulated in peripheral blood mononuclear cells of atopic asthmatics and atopic non asthmatic patients
				Brutsche, M. H. et al., Apoptosis signals in atopy and asthma measured with cDNA arrays., Clin Exp Immunol 123, 181-7. (2001).
12	7526258CD1	685389|MGC5297	1.6E-11	[Homo sapiens] Protein of unknown function, has a region of low similarity to a region of fas-activated serine threonine kinase (human FASTK), which is a serine-threonine kinase that phosphorylates RNA binding protein human TIA1 during Fas mediated apoptosis
13	7526311CD1	g1088281	7.9E-67	pyruvate dehydrogenase kinase [Homo sapiens]
				Gudi, R. et al., Diversity of the pyruvate dehydrogenase kinase gene family in humans, J. Biol. Chem. 270, 28989-28994 (1995)
13	7526311CD1	336846|PDK1	4.3E-68	[Homo sapiens][Protein kinase;Transferase;Other kinase][Cytoplasmic;Mitochondrial] Pyruvate dehydrogenase kinase 1, phosphorylates and inactivates the pyruvate dehydrogenase complex and thus regulates pyruvate metabolism
				Taylor, V. et al., 5' phospholipid phosphatase SHIP-2 causes protein kinase B inactivation and cell cycle arrest in glioblastoma cells., Mol Cell Biol 20, 6860-71 (2000).
13	7526311CD1	757382|Pdk1	2.2E-55	[Rattus norvegicus][Protein kinase;Transferase;Other kinase][Cytoplasmic;Mitochondrial] Pyruvate dehydrogenase kinase 1, phosphorylates and inactivates the pyruvate dehydrogenase complex and thus putatively regulates pyruvate metabolism
				Sugden, M. C. et al., Expression and regulation of pyruvate dehydrogenase kinase isoforms in the developing rat heart and in adulthood: role of thyroid hormone status and lipid supply, Biochem J 352, 731-8. (2000).
14	7526315CD1	g12655099	7.2E-121	Mixed lineage kinase-related kinase MRK-beta [Homo sapiens]
				Strausberg, R. L. et al. (supra)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
14	7526315CD1	476453|ZAK	3.9E-122	[Homo sapiens] Mixed lineage kinase with a leucine zipper and a sterile alpha motif, a mixed lineage kinase-like protein that stimulates the JNK/SAPK pathway and activates NF-kappaB; overexpression induces apoptosis of a hepatoma cell line
				Liu, T. C. et al., Cloning and expression of ZAK, a mixed lineage kinase-like protein containing a leucine-zipper and a sterile-alpha motif, Biochem Biophys Res Commun 274, 811-6 (2000).
14	7526315CD1	662697|Zak	2.7E-121	[Mus musculus][Protein kinase;Transferase] Mixed lineage kinase with a leucine zipper and a sterile alpha motif, activated by osmotic shock; overexpression activates the p38 (Mapk14), JNK/SAPK, ERK (Mapk3), and ERK5 (Mapk7) pathways, alpha alternative form disrupts actin stress fibers
				Gotoh, I. et al., Identification and characterization of a novel MAP kinase kinase kinase, MLTK., J Biol Chem 276, 4276-86 (2001). (supra)
15	7526442CD1	g12803641	3.5E-64	CCRK protein [Homo sapiens]
				Strausberg, R. L. et al. (supra)
15	7526442CD1	568698|CCRK	2.4E-65	[Homo sapiens][Protein kinase;Transferase] Protein containing four protein kinase domains, has a region of moderate similarity to cyclin-dependent kinase 3 (human CDK3), which is a kinase that binds to cyclin A and is required for progression from G1 to S phase
15	7526442CD1	583769|Cdk5	1.6E-22	[Mus musculus][Protein kinase;Transferase][Cell body (soma);Growth cone] Cyclin-dependent protein kinase 5, serine-threonine kinase that associates with the regulatory subunit p35 (Cdk5r) and phosphorylates neuronal proteins, involved in neuronal differentiation, regulation of myogenesis, and adaptive responses to cocaine
				Ohshima, T. et al., Targeted disruption of the cyclin-dependent kinase 5 gene results in abnormal corticogenesis, neuronal pathology and perinatal death., Proc Natl Acad Sci U S A 93, 11173-8 (1996).

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7526185CD1	157	Serine/Threonine protein kinases, catalytic domain: F24-Y157	HMMER_SMART
			KINASE PROTEIN DOMAIN TRANSFERASE PD00584: L27-G36	BLIMPS_PRODOM
			PROTEIN KINASE DOMAIN DM00004|A53714|17-262: L27-V151 DM00004|I49376|270-509: K26-G153 DM00004|P08458|20-262: I30-V151 DM00004|P38692|24-266: K26-V151	BLAST_DOMO
			Potential Phosphorylation Sites: S34, S75, S106, S137, T25, T46	MOTIFS
			Potential Glycosylation Sites: N44	MOTIFS
			Protein kinases ATP-binding region signature: I30-K53	MOTIFS
2	7526192CD1	305	Protein kinase domain: F46-G305	HMMER_PFAM
			Serine/Threonine protein kinases, catalytic domain: F46-G305	HMMER_SMART
			Eukaryotic protein kinase IPB000719: H189-L204	BLIMPS_BLOCKS
			Protein kinases signatures and profile: T173-P230	PROFILESCAN
			CASEIN KINASE I GAMMA ISOFORM CKIGAMMA TRANSFERASE SERINE/THREONINEPROTEIN ATPBINDING MULTIGENE PD026544: M1-N45	BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM00004|B56711|48-303:V48-L76 E109-R302 DM00004|A56711|46-303:V48-L76 E109-R302 DM00004|C56711|45-301:V48-L76 E109-R302 DM00004|D56406|31-276:V48-L76 E109-R302	BLAST_DOMO
			Potential Phosphorylation Sites: S19, S99, S129, S262, T84, T183, T210, T232, T247	MOTIFS
			Protein kinases ATP-binding region signature: I52-K75	MOTIFS
			Serine/Threonine protein kinases active-site signature: L193-V205	MOTIFS
3	7526193CD1	930	Signal_cleavage: M1-G68	SPSCAN
			Protein kinase domain: V46-F310	HMMER_PFAM
			Serine/Threonine protein kinases, catalytic domain: V46-K313	HMMER_SMART
			Eukaryotic protein kinase IPB000719: C168-L183, I239-G249	BLIMPS_BLOCKS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			PROTEIN REPEAT SIGNAL PRECURSOR PRION GLYCOPROTEIN NUCLEAR GPIANCHOR BRAIN MAJOR PD001091: G373-P626, G404-P626, P358-Q601, P349-Q574, P320-S519, P296-Q541	BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM00004|P38080|36-309: L52-I304 DM00004|P40494|23-287: L52-I304 DM00004|P51954|6-248: L52-I304 DM00004|P53974|23-288: L52-I304	BLAST_DOMO
			Potential Phosphorylation Sites: S7, S115, S224, S235, S311, S625, S679, S785, S815, S822, S833, S871, S879, T47, T147, T199, T221, T240, T241, T275, T389, T395, T628, T708, T743, T757, T829	MOTIFS
			Potential Glycosylation Sites: N113, N273, N667, N703, N823, N905	MOTIFS
			Serine/Threonine protein kinases active-site signature: I172-L184	MOTIFS
4	7526196CD1	118	Signal Peptide: M1-G22	HMMER
			Signal_cleavage: M1-G22	SPSCAN
			Serine/threonine dehydratase pyridoxal-phosphate attachment site IPB000634: E95-S104	BLIMPS_BLOCKS
			CYCLIN G-ASSOCIATED KINASE TRANSFERASE SERINE/THREONINEPROTEIN ATPBINDING HSGAK PD026473: M1-L40	BLAST_PRODOM
			Potential Phosphorylation Sites: S6, S21, S62, S73, S92, S113	MOTIFS
5	7526198CD1	1355	Protein kinase domain: L40-E315	HMMER_PFAM
			DnaJ molecular chaperone homology domain: E1290-S1351	HMMER_SMART
			Serine/Threonine protein kinases, catalytic domain: L40-A317	HMMER_SMART
			Eukaryotic protein kinase IPB000719: Q165-L180, I240-G250	BLIMPS_BLOCKS
			Protein kinases signatures and profile: V148-H200	PROFILESCAN
			CYCLIN G-ASSOCIATED KINASE TRANSFERASE SERINE/THREONINEPROTEIN ATPBINDING HSGAK PD039449: A317-N402	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			PROTEIN AUXILIN COAT REPEAT PHOSPHORYLATION KIAA0473 CYCLIN G-ASSOCIATED KINASE TRANSFERASE PD010124: Q1215-Q1349 PD025411: S456-V640 PD151518: L641-L1093, P868-S1235, R320-E366	BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM00004|P38080|36-309: L46-I306 DM00004|P40494|23-287: R41-I306 DM00004|P53974|23-288: R44-I306 DM00004|Q09170|169-423: R44-S305	BLAST_DOMO
			Potential Phosphorylation Sites: S6, S21, S62, S73, S93, S305, S393, S456, S530, S540, S551, S661, S726, S737, S738, S784, S811, S906, S976, S1029, S1103, S1113, S1220, S1234, S1235, S1237, S1344, T155, T186, T382, T414, T459, T611, T680, T776, T805, T949, T1118, T1156, T1165, T1244, Y412	MOTIFS
			Potential Glycosylation Sites: N677, N724, N809, N970, N1196	MOTIFS
			Serine/Threonine protein kinases active-site signature: I169-L181	MOTIFS
6	7526208CD1	490	Protein kinase domain: Y14-I252	HMMER_PFAM
			Serine/Threonine protein kinases, catalytic domain: Y14-I252	HMMER_SMART
			Eukaryotic protein kinase IPB000719: H108-L123, Y171-G181	BLIMPS_BLOCKS
			Protein kinases signatures and profile: F65-D147	PROFILESCAN
			Tyrosine kinase catalytic domain signature PR00109: H106-L124, V175-E197, V221-A243	BLIMPS_PRINTS
			KINASE PROTEIN II CALCIUM/CALMODULIN-DEPENDENT TYPE SUBUNIT CALMODULINBINDING CHAIN TRANSFERASE SERINE/THREONINEPROTEIN PD001779:I252-K303 S312-V380	BLAST_PRODOM
			KINASE PROTEIN II CALCIUM/CALMODULIN-DEPENDENT TYPE SUBUNIT CHAIN TRANSFERASE SERINE/THREONINEPROTEIN CALMODULINBINDING PD004250: E381-K469	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			PROTEIN KINASE DOMAIN DM00004|JU0270|16-262:E18-R53 V54-A243 DM00004|A44412|16-262:E18-R53 V54-A243 DM00004|P11798|15-261: E39-A243, L16-E63	BLAST_DOMO
			KINASE; DEPENDENT; II; CALMODULIN; DM05068|P11798|263-426: S244-A418	BLAST_DOMO
			Potential Phosphorylation Sites: S51, S59, S89, S312, S313, S397, T36, T47, T74, T242, T327, T328, T369	MOTIFS
			Potential Glycosylation Sites: N293, N326, N479	MOTIFS
			Protein kinases ATP-binding region signature: L20-K43	MOTIFS
			Serine/Threonine protein kinases active-site signature: V112-L124	MOTIFS
7	7526212CD1	344	Protein kinase domain: Y14-I252	HMMER_PFAM
			Serine/Threonine protein kinases, catalytic domain: Y14-I252	HMMER_SMART
			Eukaryotic protein kinase IPB000719: H108-L123, Y171-G181	BLIMPS_BLOCKS
			Protein kinases signatures and profile: F65-D147	PROFILESCAN
			Tyrosine kinase catalytic domain signature PR00109: H106-L124, V175-E197, V221-A243	BLIMPS_PRINTS
			KINASE PROTEIN II CALCIUM/CALMODULIN-DEPENDENT TYPE SUBUNIT CALMODULINBINDING CHAIN TRANSFERASE SERINE/THREONINEPROTEIN PD001779: I252-K324	BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM00004|JU0270|16-262:E18-R53 V54-A243 DM00004|A44412|16-262:E18-R53 V54-A243 DM00004|P08414|44-285: E19-T242 DM00004|P11798|15-261: E39-A243, L16-E63	BLAST_DOMO
			Potential Phosphorylation Sites: S51, S59, S89, T36, T47, T74, T242, T316, T317	MOTIFS
			Potential Glycosylation Sites: N293, N315	MOTIFS
			Protein kinases ATP-binding region signature: L20-K43	MOTIFS
			Serine/Threonine protein kinases active-site signature: V112-L124	MOTIFS
8	7526213CD1	89	Potential Phosphorylation Sites: S5, S56, S80, T52	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Hexokinase family IPB001312: S10-G24	BLIMPS_BLOCKS
9	7526214CD1	88	Potential Phosphorylation Sites: S5, S56, S67, T52	MOTIFS
			Hexokinase family IPB001312: S10-G24	BLIMPS_BLOCKS
10	7526228CD1	137	Signal_cleavage: M1-A15	SPSCAN
			PROTEIN KINASE DOMAIN DM00004|I38044|100-349: V38-A117 DM00004|P08630|329-573: E35-N114 DM00004|Q08881|361-604: E35-L112	BLAST_DOMO
			Potential Phosphorylation Sites: S14, S67, S69	MOTIFS
			Leucine zipper pattern: L112-L133	MOTIFS
			Protein kinases ATP-binding region signature: V42-K63	MOTIFS
11	7526246CD1	243	Regulator of G protein signaling domain: T54-C175	HMMER_PFAM
			Regulator of G protein signalling domain: T54-C175	HMMER_SMART
			GPCR kinase signature PR00717: F171-N183	BLIMPS_PRINTS
			Regulator of G protein signalling domain proteins PF00615: M15-K21, F162-K178	BLIMPS_PFAM
			RECEPTOR KINASE TRANSFERASE SERINE/THREONINEPROTEIN ATPBINDING BETAADRENERGIC COUPLED PROTEIN MULTIGENE FAMILY PD007430: M1-V53	BLAST_PRODOM
			KINASE; THREONINE; ATP; SERINE; DM01747|P21146|152-191: E152-S187	BLAST_DOMO
			N-TERMINAL DOMAIN DM05135|P21146|33-150: L33-E151 DM05135|P32865|33-150: L34-E151 DM05135|Q09639|34-149: L34-I150	BLAST_DOMO
			Potential Phosphorylation Sites: S29, S38, S60, S127, S168, T97	MOTIFS
			Cell attachment sequence: R158-D160	MOTIFS
12	7526258CD1	463	CELL CYCLE PROGRESSION PROTEIN FAST KINASE PD041692: L200-P417	BLAST_PRODOM
			FAST KINASE PD135789: M1-R201	BLAST_PRODOM
			Potential Phosphorylation Sites: S94, S246, S332, S373, S441, T138, T336, T365	MOTIFS
13	7526311CD1	184	Signal Peptide: M1-G18, M1-A21	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Signal_cleavage: M1-A21	SPSCAN
			Cytosolic domain: K163-T184 Transmembrane domain: W143-W162 Non-cytosolic domain: M1-T142	TMHMMER
			KINASE DEHYDROGENASE TRANSFERASE PD01976: P54-G66, N69-S117	BLIMPS_PRODOM
			KINASE PYRUVATE DEHYDROGENASE TRANSFERASE DEHYDROGENASE-LIPOAMIDE MITOCHONDRIAL PRECURSOR TRANSIT PEPTIDE MITOCHONDRION PD004994: V42-I135	BLAST_PRODOM
			PYRUVATE DEHYDROGENASE-LIPOAMIDE KINASE ISOZYME 1, MITOCHONDRIAL PRECURSOR EC 2.7.1.99 DEHYDROGENASE ISOFORM 1 TRANSFERASE TRANSIT PEPTIDE MITOCHONDRION MULTIGENE FAMILY PD174825: M1-E39	BLAST_PRODOM
			KINASE; DEHYDROGENASE; PYRUVATE; ACID; DM01978|A55305|2-103: A37-E130 DM01978|I55465|28-129: F28-E130 DM01978|I70159|2-103: A37-E130 DM01978|I70160|1-99: V42-E130	BLAST_DOMO
			Potential Phosphorylation Sites: S38, S58, S117, S128, S170	MOTIFS
14	7526315CD1	386	Protein kinase domain: L16-V266	HMMER_PFAM
			Serine/Threonine protein kinases, catalytic domain: L16-L262	HMMER_SMART
			Protein kinases signatures and profile: I107-T161	PROFILESCAN
			PROTEIN KINASE DOMAIN DM00004|A53800|119-368: E20-K221 DM00004|A55318|159-389: D15-W216 DM00004|JC2363|126-356: D15-W216 DM00004|Q05609|553-797: E20-S233	BLAST_DOMO
			Potential Phosphorylation Sites: S61, S89, S96, S233, S273, S277, S295, S341, S346, S360, S365, T345, Y274	MOTIFS
			Potential Glycosylation Sites: N97, N159, N340	MOTIFS
			Leucine zipper pattern: L225-L246, L232-L253	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Serine/Threonine protein kinases active-site signature: V129-I141	MOTIFS
15	7526442CD1	152	Eukaryotic protein kinase IPB000719: H119-Q134	BLIMPS_BLOCKS
			PROTEIN KINASE DOMAIN DM00004|I49592|6-276: L7-R131 DM00004|P23437|6-286: R9-R131 DM00004|P29620|21-289: I10-P130 DM00004|Q02399|6-276: L7-R131	BLAST_DOMO
			Protein kinases ATP-binding region signature: I10-K33	MOTIFS

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
16/ 7526185CB1/ 4430	2471-4430, 1476-1534, 1-754, 1679-1715	FL1002225_382	1	4430
		95083778J1	313	1222
		GBI.NT_009952_014.10.edit2	358	592
		GBI.NT_009952_014.10.edit1	649	4430
		72678960V1	938	1574
		72678288V1	1084	1754
		g9777972	1095	1765
		72682030V1	1185	1978
		g14503665	1198	1884
		g11642692	1204	1880
		73197364D1	1208	1859
		g14810994	1214	2029
		72680814V1	1223	2017
		73197252V1	1255	1917
		g24471308	1255	2004
		73199082D1	1259	1957
		73197393D1	1282	2255
		73196694D1	1298	1950
		g12769183	1321	2017
		g23286620	1335	2004
		g23286086	1352	2002
		g29389943	1354	2004
		g21980207	1383	2002
		g12763752	1396	2064
		g13531552	1469	2171
		g31267289	1503	2297
		g13341861	1506	2183
		g11643902	1515	2243
		g31271373	1528	2457
		g31069857	1546	2439
		g30307375	1560	2428
		g30307376	1560	2458
		g13534533	1582	2282
		8568096T1 (KIDNFEC01)	1582	2429
		g19119842	1587	2305
		g16200364	1599	2338
		g14814256	1605	2350
		g30463287	1625	2433
		8567187T1 (KIDNFEC01)	1638	2413
		g30758954	1640	2392
		g31295054	1666	2445
		71382643V1	1693	2395
		g24809844	1716	2469
		g24806937	1718	2467
		g29391117	1720	2469

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		8557426T1 (LUNGNOT30)	1724	2451
		8628262H1 (UTREDMF02)	1724	2463
		8628262J1 (UTREDMF02)	1724	2467
		g12760412	1772	2468
		g10745442	1775	2468
		g14294389	1782	2468
		g24794954	1782	2469
		7623349H1 (HEARFEE03)	1783	2431
		g23288713	1786	2468
		g23295470	1801	2467
		8215062H1 (FIBRTXC01)	1809	2468
		g23293825	1816	2469
		g24810755	1829	2469
		g21981399	1833	2469
		g31148999	1852	2469
		g19751033	1858	2469
		g22697148	2196	2847
		g12877899	2386	3249
		g11261005	3402	4093
		4289337F6 (BRABDIR01)	3471	4274
		g24795218	3698	4426
17/ 7526192CB1/ 3276	1999-3276, 910-1003, 1-224, 1546-1612	GBI_NT_011255_001.13.edit1	1	3276
		g14077475	734	1407
		6981630H1 (BRAIFER05)	1223	1711
		6306286F7 (NERDTDN03)	1860	2512
		6306286F8 (NERDTDN03)	1860	2524
		6306286T6 (NERDTDN03)	1925	2468
		55139024H1	2447	3057
		55139140J1	2552	3009
18/ 7526193CB1/ 3910	3709-3732, 2091-3219, 1-823, 3788-3910	7217965H1 (COLNTMC01)	1	344
		7266654H2 (NOSEDIC01)	20	367
		GBI.938794.82	20	3910
		g19373027	99	612
		g6992730	340	775
		3617328F6 (EPIPNOT01)	464	775
		55139719H1	486	797
		1328791H1 (PANCNOT07)	500	746
		g9707306	520	779
		g4740126	520	779
		g9705742	520	779
		g4987918	520	779
		g5637149	520	779
		3023352H1 (PROSDIN01)	537	799

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		1679654H1 (STOMFET01)	552	768
		g5849814	560	779
		6844123H1 (KIDNTMN03)	567	790
		55099275J1	582	1094
		8018416F6 (BMARTXE01)	587	1213
		3574386H1 (BRONNOT01)	610	895
		72717764V1	616	1173
		72719264V1	634	1307
		55099283J1	698	1094
		g10091676	706	1162
		55139943H1	776	1097
		55139835J1	778	1097
		55139827J1	797	1101
		55139819J1	799	1093
		g13410426	805	1348
		g9331447	824	1456
		g10734144	824	1504
		g9336431	825	1466
		g10143324	830	1503
		g7930221	834	1412
		g11316872	840	1124
		g7254240	840	1426
		55139803J1	849	1097
		55139811J1	868	1096
		g6868467	885	1348
		g20968062	1048	1744
		g20856405	1115	1714
		g14254976	1124	1734
		g14254972	1127	1742
		g20967828	1146	1711
		g8039829	1217	1481
		g8039887	1222	1524
		g14404428	1246	1525
		7275351H2 (LIVRUNE01)	1249	1757
		g14404429	1251	1509
		g21012570	1295	1751
		g14451882	1321	1832
		g14453118	1334	1889
		4291033H1 (BRABDIR01)	1376	1647
		72337884V1	1376	1670
		72337654V1	1376	1768
		72336903V1	1376	1768
		72338409V1	1376	1768
		72337087V1	1376	1768
		72338101V1	1376	1768
		72337058V1	1376	1768
		72338322V1	1376	1768

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		72338656V1	1376	1768
		72337641V1	1376	1768
		72338238V1	1376	1768
		72337228V1	1376	1768
		72338353V1	1376	1768
		72766921V1	1376	1768
		72338557V1	1376	1768
		72338434V1	1376	1768
		72336974V1	1376	1768
		72338470V1	1376	1768
		72338136V1	1376	1768
		72338013V1	1376	1768
		4291033F6 (BRABDIR01)	1376	1768
		72338336V1	1376	1768
		72337535V1	1376	1768
		72338790V1	1376	1768
		72338126V1	1376	1768
		72338444V1	1376	1768
		72338450V1	1376	1768
		72337857V1	1376	1768
		72337183V1	1387	1768
		g21012239	1466	1742
		g12199215	1470	1829
		g14404390	1491	2092
		g28118293	1497	1880
		7993480H1 (UTRSDIC01)	1502	1820
		8018416R6 (BMARTXE01)	1520	2235
		7067749H1 (BRATNOR01)	1599	2231
		6345837H1 (LUNGDIS03)	1961	1989
		6345837H1 (LUNGDIS03)	1992	2298
		6038785H1 (PITUNOT06)	2014	2654
		g2354017	2076	2336
		8360236J1 (MIXDUNN06)	2092	2727
		g12371898	2135	2425
		g12361664	2135	2448
		5781301F6 (BRAXNOT03)	2147	2580
		g12361674	2153	2443
		56057236H1	2226	2705
		g12370692	2476	2624
		g10460662	2508	2889
		5497716F6 (BRABDIR01)	2511	2969
		g12233185	2530	2930
		g14345747	2807	3117
		6327887H1 (BRANDIN01)	2862	3393
		5497716R6 (BRABDIR01)	2870	3085
		6245571H1 (TESTNOT17)	3117	3543
		5772648H1 (BRAINOT20)	3343	3863

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		g2237352	3458	3789
		5790314H1 (FIBRTXS07)	3569	3807
		5786172H1 (FIBRTXS07)	3569	3807
		5786527H1 (FIBRTXS07)	3569	3807
19/ 7526196CB1/ 4380	464-663, 2075-2164, 1-349, 3320-4380	GBI.g29789976.edit1	1	4380
		g9772401	41	683
		9505159U1	763	1641
		9524857U1	818	1676
		9649412U2	884	1835
		9505172U1	887	1743
		9524956U1	887	1834
		9649412U1	887	1835
		9611509U1	897	1740
		9611509U3	900	1709
		9509648U1	954	1835
		7754868J1 (SPLNTUE01)	1032	1590
		9580055U3	1035	1895
		9600429U1	1036	1729
		9600055U1	1036	1835
		9600429U3	1039	1835
		9580055U1	1065	1676
		55095641J1	1141	1738
		72484222D1	1169	1738
		72481795D1	1176	1738
		72616020V1	1188	1738
		72484068D1	1189	1737
		8516684H1 (HNT2TXF01)	1198	1920
		8757725H1 (TLYJTXN01)	1198	2052
		72481336D1	1202	1738
		6831090J1 (SINTNOR01)	1304	1944
		7719693H1 (SINTFEE02)	1353	2006
		6819080J1 (BRAUNOR01)	1354	1879
		8016820J1 (BMARTXE01)	1354	1985
		8021257J1 (BMARTXE01)	1354	2016
		8016427J1 (BMARTXE01)	1354	2051
		7050749F6 (BRACNOK02)	1354	2086
		7600819J1 (ESOGTME01)	1356	2011
		55144872J1	1411	2084
		55144879J1	1411	2094
		g11295923	1498	2168
		90217915J1	2279	2898
		90218047J1	2279	2901
		90217923J1	2279	2918
		90218039J1	2279	2945
		90217907J1	2279	2987

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		90217931J1	2279	3041
		90217947J1	2279	3091
		90218031J1	2279	3111
		90218023J1	2279	3151
		9679740U1	2285	3136
		95103533H1	2361	2898
		7208870H1 (FIBPFEA01)	2365	3037
		7050749R6 (BRACNOK02)	2368	3037
		822846R1 (KERANOT02)	2392	2942
		7253472H1 (BRAMNOA01)	2398	3036
		7703764H1 (UTRETUE01)	2406	2999
		g12097556	2422	3076
		7050236H1 (BRACNOK02)	2429	3046
		7115279H1 (BRAENOK01)	2430	3037
		g19212752	2447	3014
		7685360H1 (BRABDIK02)	2500	3037
		6946127F6 (FTUBTUR01)	2538	3118
		90218039J1	3013	3077
		g30442850	3108	4157
20/ 7526198CB1/ 4293	1-48, 3480-3610, 1319-1887	73232879V1	1	622
		73232879D1	1	624
		GBI.g29789976.edit1	1	4163
		g14002261	3	677
		g10937540	49	748
		g13997818	76	607
		8684117H1 (BRAIUNF01)	83	981
		g22275488	86	731
		g22660086	86	826
		8042207H1 (OVARTUE01)	101	650
		7751875H1 (HEAONOE01)	102	660
		7441147H1 (ADRETUE02)	102	693
		g30216686	138	660
		9713909U2	138	847
		90049479J1	139	730
		90049355F6	139	807
		90049387J1	139	833
		9713909U1	139	847
		90049363J1	139	848
		90049379H1	139	914
		90049495J1	139	926
		90049355H1	139	1002
		60215662U1	194	827
		g21012603	299	882
		g21012604	318	882
		9679740U1	2082	2933

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		90217915J1	2092	2695
		90218047J1	2092	2698
		90217923J1	2092	2715
		90218039J1	2092	2742
		90217907J1	2092	2784
		90217931J1	2092	2838
		90217947J1	2092	2888
		90218031J1	2092	2908
		90218023J1	2092	2948
		95103533H1	2158	2695
		7208870H1 (FIBPFEA01)	2162	2834
		7050749R6 (BRACNOK02)	2165	2834
		822846R1 (KERANOT02)	2189	2739
		7253472H1 (BRAMNOA01)	2195	2833
		7703764H1 (UTRETUE01)	2203	2796
		g12097556	2219	2873
		7050236H1 (BRACNOK02)	2226	2843
		7115279H1 (BRAENOK01)	2227	2834
		g19212752	2244	2811
		7685360H1 (BRABDIK02)	2297	2834
		6946127F6 (FTUBTUR01)	2335	2915
		90218039J1	2810	2874
		g30442850	2905	3964
		GBI.g29789976.edit2	3908	4293
21/ 7526208CB1/ 6538	2067-2098, 4319-5404, 4173-4201, 1-1485, 6070-6538, 2539-3468	2944771F7 (BRAITUT23)	1	581
		GBI_NT_016354_004.13.edit1	21	6538
		8018737J1 (BMARTXE01)	438	1053
		g3422499	581	1057
		g3330808	589	1057
		8198864H1 (BRAINOR03)	906	1436
		8198864J1 (BRAINOR03)	1332	1978
		7580306H1 (BRAIFEC01)	1483	2045
		8159951J1 (MIXDTME02)	1486	2135
		8161727H1 (MIXDTME02)	1645	2142
		90041465H1	2127	2864
		90041465F6	2129	2864
		g5926184	2304	2751
		8267027H1 (MIXDUNF02)	2382	2971
		8498563H1 (BRSMTXF01)	2543	3089
		g2162976	2564	3118
		8123676H1 (HEAONOC01)	2568	3068
		71891427V1	3104	3725

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		71893456V1	3255	3883
		2228879F6 (PROSNOT16)	3376	3897
		71892008V1	3450	4136
		g11290981	3459	4057
		1236920F1 (LUNGFET03)	3476	3968
		6517912F8 (BRAFTDT02)	3478	4027
		g11976492	3479	4046
		g24786219	3481	4172
		56010211H1	3489	4296
		g10823987	3502	4173
		g24787165	3522	4172
		g30853338	3533	4165
		g18521727	3555	4177
		g13285388	3564	4310
		g6890090	3578	4049
		3967421F7 (PROSTUT10)	3643	4157
		3967421F6 (PROSTUT10)	3643	4275
		g2836991	3652	4172
		g28094190	3663	4244
		1803939F6 (SINTNOT13)	3676	4334
		g22767266	3699	4147
		g10825536	3806	4475
		4827574F6 (BLADDIT01)	3959	4517
		4827574T8 (BLADDIT01)	3972	4672
		7682581T8 (BRABDIK02)	4006	4691
		g24779423	4007	4745
		g24794912	4028	4745
		g12102435	4040	4746
		g9876920	4043	4695
		g19734318	4067	4745
		g23285997	4076	4745
		g27792216	4078	4745
		g11364160	4104	4659
		g24775063	4116	4745
		g19729780	4155	4745
		g19755171	4165	4745
		g21175144	4184	4745
		1803939T6 (SINTNOT13)	4194	4686
		1625628T6 (COLNPOT01)	4195	4685
		2228879T6 (PROSNOT16)	4215	4689
		4827574T6 (BLADDIT01)	4216	4688
		g23286214	4242	4745
		2851843T6 (BRSTTUT13)	4251	4702
		2153236F6 (BRAINOT09)	4284	4745
		7621418J1 (HEARFEE03)	4873	5519
		g12413646	4878	5586
		g10319848	5089	5766

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		6610468H2 (KIDNTMC01)	5144	5808
		414471F1 (BRSTNOT01)	5483	6067
		6480671H1 (PROSTMC01)	5491	6053
		7326473R8 (THYMNOE02)	5502	6329
		g10318203	5571	6067
		7621418H1 (HEARFEE03)	5580	6040
		g3099121	5586	6065
		7012818F7 (KIDNNOC01)	5587	6067
		7012918F8 (KIDNNOC01)	5587	6069
		g10035211	5603	6069
		g4188696	5614	6067
22/ 7526212CB1/ 2349	1442-1485, 2116-2349, 1-1049	2944771F7 (BRAITUT23)	1	581
		GBI_NT_016354_004.13.edit1	1	2290
		8018737J1 (BMARTXE01)	438	1053
		g3422499	581	1057
		g3330808	589	1057
		8198864H1 (BRAINOR03)	906	1436
		8198864J1 (BRAINOR03)	1332	1978
		7580306H1 (BRAIFEC01)	1483	2045
		73073134V1	1608	2349
23/ 7526213CB1/ 8015	6365-6398, 7104-7128, 4028-5749, 1-2254, 2817-3243	GBI_NT_016354_003.15.edit1	1	8009
		90214127H1	203	877
		9815193U2	203	877
		9775316U2	203	936
		9822048U1	571	1314
		9770976U2	571	1436
		9785981U1	571	1450
		9770976U1	577	1378
		9785972U1	586	1449
		9822048U2	594	1332
		9746466U2	652	1450
		9773732U2	731	1652
		9784110U2	747	1525
		9784110U1	756	1532
		9770966U1	799	1670
		9796042U2	806	1686
		9746439U1	833	1552
		9746180U2	833	1695
		9746439U2	833	1721
		9738822U2	833	1768
		9770980U1	843	1675

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		9770980U2	849	1612
		9746180U1	850	1644
		9770984U2	850	1675
		9770970U1	855	1615
		9770970U2	856	1771
		9770962U2	857	1624
		9746239U2	857	1678
		9746359U1	915	1764
		9811817U2	928	1675
		9822051U2	928	1768
		9822051U1	928	1899
		9770964U1	933	1753
		9770964U2	933	1779
		9773790U2	934	1774
		9785982U2	943	1759
		9746294U1	947	1898
		9746294U2	950	1844
		9822053U1	1827	2719
		9785984U2	1829	2642
		9785975U2	1829	2704
		9785975U1	1829	2737
		9785984U1	1830	2720
		9746440U2	2108	3060
		9746414U1	2154	3036
		9770983U2	2155	3043
		9786418U1	2160	3084
		9746414U2	2161	3052
		9770973U2	2172	3054
		9822054U1	2174	3015
		9822054U2	2180	2972
		9785985U1	2212	3066
		9785976U1	2212	3163
		9785976U2	2215	2958
		9785985U2	2215	3030
		9770975U1	2277	3051
		9770979U1	2281	3068
		9785986U2	2422	3338
		9822055U1	2430	3130
		9785986U1	2430	3212
		9785977U2	2430	3244
		9822055U2	2430	3322
		9784111U2	2434	3184
		g10934151	5975	6837
		g15763429	6557	7314
		g18513541	7215	7903
		g21477809	7241	8015

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
24/ 7526214CB1/ 7945	7039-7063, 1-2933, 3938-5684, 6300-6373	GBI_NT_016354_003.15.edit1	1	7945
		9775320U2	203	668
		9775320U1	203	955
		9785981U2	492	1352
		9785990U2	507	1304
		9770976U2	507	1371
		9785981U1	507	1385
		9770976U1	513	1313
		9785972U2	514	1304
		9785972U1	522	1384
		9746466U2	588	1385
		9770978U2	692	1396
		9775320U2	707	787
		9770974U1	735	1548
		9796042U2	742	1621
		9746439U1	769	1487
		9746439U2	769	1656
		9738822U2	769	1703
		9746215U2	785	1551
		9770984U2	786	1610
		9770970U1	787	1550
		9770962U2	787	1559
		9770970U2	788	1706
		9770968U2	790	1712
		9746215U1	791	1488
		9746239U2	792	1613
		9770968U1	796	1676
		9811817U2	863	1610
		9822051U2	863	1703
		9822051U1	863	1834
		9770964U1	868	1688
		9770964U2	868	1714
		9773790U2	869	1709
		9785982U2	878	1694
		9746294U1	882	1833
		9746294U2	885	1779
		9822053U1	1762	2657
		9785984U2	1764	2579
		9785975U2	1764	2642
		9785975U1	1764	2675
		9785984U1	1765	2658
		9746295U1	2418	3318
		9811820U2	2583	3356
		95037369J1	2664	3334

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		95048903H1	2664	3380
		95049151H1	2664	3382
		95048851H1	2664	3382
		95037269J1	2664	3382
		95049119J1	2664	3389
		95049103J1	2664	3405
		95049127J1	2664	3422
		95048935H1	2664	3431
		95049011H1	2664	3445
		95049135J1	2664	3454
		95048927H1	2664	3458
		95048943H1	2664	3458
		95037385H1	2664	3459
		95037377J1	2664	3464
		95049183H1	2664	3469
		95037277J1	2664	3469
		95049191H1	2664	3470
		95037393H1	2664	3471
		95048891H1	2664	3480
		95048975H1	2664	3510
		95048983H1	2664	3511
		95037293H1	2664	3532
		95048911H1	2664	3562
		95048919J1	2668	3458
		95049075H1	2668	3609
		95049011J1	2668	3617
		95048991H1	2706	3617
		95048935J1	2711	3617
		95037293J1	2776	3617
		95037369H1	2779	3617
		95037385J1	2779	3617
		95048951J1	2779	3617
		9743770U1	2781	3654
		95048903J1	2782	3617
		90214227J1	2784	3654
		95037377H1	2804	3617
		9775731U2	2810	3654
		90214259J1	2823	3653
		9775730U2	2848	3654
		9785996U2	2849	3628
		9785996U1	2849	3651
		9785978U1	2849	3654
		9775731U1	2854	3654
		90214227R6	2876	3653
		9775722U2	2876	3654
		9775718U2	2876	3654
		9775723U2	2899	3650

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		9775725U2	2906	3654
		9775719U1	2915	3654
		9791941U1	2949	3654
		9775718U1	2952	3654
		9775720U1	2971	3654
		g24792102	3008	3682
		g14047307	3181	3922
		8138972T1 (SPLNNOT10)	3471	4320
		g19732719	3797	4551
		8760604H1 (MYEPUNN01)	4066	4870
		8720327H1 (TLYJUNF01)	4320	5134
		8717314H1 (TLYJTXF03)	4541	5355
		8502029H1 (KIDEUNF01)	4764	5490
		g21170624	5247	5934
		g10934151	5910	6772
		g15763429	6492	7250
		g10153702	6854	7555
		g18513541	7150	7839
		g21477809	7176	7945
25/ 7526228CB1/ 3149	1298-1355, 2272-3149	g14083204	1	528
		8507486H1 (SMCCTXF01)	1	719
		GBI_NT_007299_017.12.edit1	30	3149
		7953010H1 (SYNONOC01)	217	706
		73414963V1	229	740
		g15748947	239	1013
		8711164H1 (MYEPUNF01)	243	910
		g15759491	263	996
		95104290J1	402	1249
		95104234H1	1515	2354
		71634372V1	1538	2050
		95104302H1	1569	2354
		8208729H1 (LIVRTXS02)	1889	2602
		g13521551	2023	2773
		71769306V1	2112	2648
		71638426V1	2124	2819
		71635816V1	2147	2671
		g12365044	2211	2688
		g11014334	2566	3119
26/ 7526246CB1/ 3617	1-1020, 2999-3617	95079085H1	1	563
		GBI.1859885.71	118	3617
		6828803J1 (SINTNOR01)	407	1087
		g14290834	412	1145
		7401136H1 (SINIDME01)	506	1090

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		7741107H1 (THYMNOE01)	521	563
		g15763036	703	1270
		95079085H1	869	1158
		7741107H1 (THYMNOE01)	869	1501
		g11937225	1194	1870
		g13915792	1194	1985
		g13916980	1194	1992
		9743772U2	1358	2314
		8183605H1 (EYERNON01)	1371	2039
		95078961H1	1373	2314
		95079093H1	1376	2303
		95078713J1	1385	2314
		g19891326	1421	2000
		g19895139	1423	2016
		9817121U4	1423	2314
		g19370851	1429	2101
		56082957H1	1439	2254
		9817121U3	1455	2314
		g14054691	1457	2209
		95078937J1	1466	2314
		95078777H1	1471	2313
		95078913J1	1471	2314
		95078729J1	1478	2314
		95078985H1	1480	2314
		95043360J1	1480	2314
		95078993H1	1492	2313
		95079077H1	1494	2313
		95043452J1	1494	2314
		90155990J1	1495	2351
		95078837J1	1497	2314
		9743371U1	1497	2314
		90155858J1	1497	2351
		g12766544	1509	2150
		90155874J1	1509	2351
		95079005F6	1515	2314
		95078793H1	1519	2303
		95079005H1	1520	2314
		g11641958	1521	2249
		95043436J1	1529	2314
		9805472U2	1529	2314
		g8658271	1539	2152
		90151356J1	1542	2130
		90161465H1	1542	2336
		90161565H1	1542	2367
		90161481H1	1542	2412
		9805472U1	1544	2314
		g12612323	1545	2247

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		7721181J1 (THYRDIE01)	1557	2187
		95078893H1	1568	2313
		95078861H1	1568	2314
		95078885H1	1568	2314
		72677129V1	1592	2177
		95079093F6	1594	2303
		95079013J1	1598	2314
		g12684079	1599	2282
		90155866J1	1599	2351
		g18515296	1601	2189
		9743772U1	1613	2314
		90154817H1	1616	2267
		95078705H1	1616	2314
		95078829J1	1616	2314
		g22371114	1625	2252
		g21120536	1631	2190
		95078853J1	1638	2313
		g16175992	1648	2404
		7726720J1 (UTRCDIE01)	1653	2261
		7690765J1 (PROSTME06)	1656	2242
		g19367868	1661	2344
		72681393V1	1677	2397
		g15348804	1679	2352
		993893R6 (COLNNOT11)	1688	2368
		g10454044	1714	2320
		g9720208	1715	2528
		7621966J1 (HEARFEE03)	1728	2280
		6054035F6 (BRAENOT04)	1737	2340
		g30775472	1740	2392
		95079029J1	1744	2314
		g10320808	1773	2465
		90155982H1	1782	2351
		g14814460	1797	2414
		7723759J1 (THYRDIE01)	1824	2428
		8752975H1 (TLYJTXN02)	2599	3242
27/ 7526258CB1/ 1955	1-41, 1882-1955, 1529-1598	GBI.NT_007914_013.10.edit1	1	1955
		g13910802	81	946
		73381212D1	715	1172
		g19374315	812	1528
		7212021H2 (BLYRTXT03)	987	1540
28/ 7526311CB1/ 2937	1394-1429, 1-92, 2612-2937	GBI.928308.PT127_1	1	2937
		95117349H1	122	658
		95117218H1	122	690

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		95117318H1	122	739
		GBI.95117318CL1	123	1393
		95117318J1	539	1393
		95117349J1	621	1393
		95117218J1	807	1393
		g10357922	1099	1644
		6701408H1 (DRGCNOT02)	1237	1848
		2745158H1 (LUNGTUT11)	1251	1561
		8374138J1 (MIXDUNN16)	1390	2042
		g10208513	1392	2069
		g11112201	1396	2052
		g31804533	1430	1987
		7651467F6 (STOMTDE01)	1495	1949
		7651467H1 (STOMTDE01)	1502	1940
		g12336360	1531	2266
		55094066H1	1542	2039
		55094066J1	1542	2039
		2927987F6 (TLYMNOT04)	1549	2106
		g24902268	1579	1965
		g31010996	1658	2132
		g9772813	1731	2380
		g12427833	1737	2490
		g13581931	1759	2540
		6280867T8 (SKINDIA01)	1837	2534
		g23285666	1853	2606
		g8909555	1871	2301
		1649261F6 (PROSTUT09)	1880	2223
		2596906F6 (OVARTUT02)	1925	2496
		268900T6 (HNT2NOT01)	1960	2513
		2921276F6 (SININOT04)	1971	2483
		g1476946	1984	2421
		g1479766	1984	2491
		55005237J1 (PHDEDNV02)	1984	2567
		g1484668	1984	2611
		g1927463	1992	2496
		907626R2 (COLNNOT09)	2000	2404
		g11008053	2015	2565
		2921293T6 (SININOT04)	2021	2502
		g1505878	2033	2567
		770052R1 (COLNCRT01)	2038	2544
		g434511	2060	2394
		8450642J1 (MIXDTUN01)	2062	2561
		g11014883	2066	2606
		1649261T6 (PROSTUT09)	2083	2510
		g10821187	2107	2572
		8398230T1 (SPLNNOT04)	2109	2500
		7628312H1 (GBLADIE01)	2129	2573

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		g2706188	2136	2565
		g3899217	2139	2582
		g5742157	2147	2567
		4163995T6 (BRSTNOT32)	2154	2547
		g7319568	2156	2606
		g4188364	2160	2609
		6562933H1 (MCLDTXT04)	2169	2741
		g11512426	2182	2609
		g274460	2223	2552
		g2567099	2240	2586
		g7319495	2246	2606
		g4620813	2247	2568
		g13584179	2247	2602
		2927987T6 (TLYMNOT04)	2328	2878
		8556290T2 (LUNGNOT30)	2355	2860
		1299477T6 (BRSTNOT07)	2374	2890
		g19758531	2472	2913
		g10810599	2478	2937
29/ 7526315CB1/ 6122	1-88, 983-2443, 3020-5194	GBI.928294.PT122_0	1	6121
		9790480U1	193	871
		9709180U2	193	871
		9807280U2	193	871
		9807280U1	193	871
		9709180U1	193	871
		71866765V1	310	931
		72697902V1	573	982
		72343088V1	867	1481
		72343412V1	867	1492
		72343082V1	867	1496
		72343409V1	867	1573
		72343152V1	867	1595
		72343264V1	890	1609
		72342802V1	919	1643
		72343571V1	922	1655
		72006034V1	1103	1948
		8514925H1 (BRSTUNF01)	2143	2950
		8699971H1 (LIVRTXF01)	2339	3054
		6610909J1 (PLACFER06)	2423	3023
		6935902F8 (SINTTMR02)	2471	3073
		7314251H1 (UTREDME02)	2669	3237
		73396792V1	2738	3490
		73396412V1	2738	3554
		73396412D1	2738	3554
		8286949F6 (OVARDIN02)	2800	3574
		8286949T6 (OVARDIN02)	3049	3795

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		7625836H1 (KIDNFEE02)	3247	3955
		g12613318	4679	5357
		1501689F6 (SINTBST01)	4789	5413
		g12095619	5055	5683
		8243847J1 (BONEUNR01)	5121	5809
		g11979244	5167	5854
		3016112T6 (MUSCNOT07)	5317	6012
		8546927T1 (OVARTUT01)	5343	6045
		g12763553	5356	6064
		g12758671	5370	6043
		8736604J1 (BRAJNON03)	5370	6116
		7753003J1 (HEAONOE01)	5405	6106
		58004288J1	5410	6122
		g23283197	5424	6117
		7752327H1 (HEAONOE01)	5458	6108
		58004372H1	5461	6122
		896404T2 (BRSTNOT05)	5479	6079
		g24799829	5493	6117
		1682961T7 (PROSNOT15)	5495	6063
		g15996499	5524	6117
30/ 7526442CB1/ 1914	1826-1914	g30290081	1	437
		90004721J1	1	647
		GBI.g22046009.edit1	1	1893
		GBI.FL931374.29	1	1914
		90134244J1	260	862
		90134244H1	260	863
		70623209V1	684	1206
		71565564V1	684	1379
		70626918V1	720	1262
		71563390V1	741	1452
		71564044V1	744	1453
		g1728583	748	1386
		71567352V1	750	1276
		70623897V1	760	1283
		70622825V1	762	1262
		70645387V1	784	1394
		70643337V1	789	1234
		71566273V1	789	1378
		70494443V1	791	1267
		71567483V1	794	1261
		90004721H1	796	1623
		7925523H2 (COLNTUS02)	807	1405
		70625410V1	814	1325
		70625722V1	818	1406
		90004737H1	821	1623

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		70622349V1	845	1278
		71564547V1	845	1332
		71565477V1	856	1377
		6853128H1 (BRAIFEN08)	863	1451
		6200676H1 (PITUNON01)	876	1423
		6200476H1 (PITUNON01)	880	1464
		70622609V1	953	1403
		8587450T1 (SCOMDIC01)	960	1772
		g3894478	961	1395
		71563494V1	969	1452
		6537489H1 (OVARDIN02)	1474	1914

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
16	7526185CB1	UTREDMF02
17	7526192CB1	NERDTDN03
18	7526193CB1	BRABDIR01
19	7526196CB1	BRACNOK02
20	7526198CB1	BRACNOK02
21	7526208CB1	BLADDIT01
22	7526212CB1	BRAINOR03
24	7526214CB1	MYEPUNN01
25	7526228CB1	MYEPUNF01
26	7526246CB1	THYMNOE01
27	7526258CB1	BLYRTXT03
28	7526311CB1	SININOT04
29	7526315CB1	OVARDIN02
30	7526442CB1	PITUNON01

Table 6

Library	Vector	Library Description
BLADDIT01	pINCY	Library was constructed using RNA isolated from diseased bladder tissue removed from a 73-year-old male during a total cystectomy. Pathology indicated the bladder mucosa showed mild chronic cystitis. Pathology for the associated tumor tissue indicated invasive grade 3 adenocarcinoma, which formed a friable mass situated within the proximal urethra, 14 cm from the distal urethral resection margin. The tumor invaded superficially into, but not through, muscularis propria.
BLYRTXT03	pINCY	Library was constructed using RNA isolated from a treated Raji cell line derived from the B-lymphocyte cells of an 11-year-old Black male (ATCC CCL-86). The cells were treated for 18 hours with 10ng/ml of interleukin 18 (IL-18). Pathology indicated Burkitt's lymphoma.
BRABDIR01	pINCY	Library was constructed using RNA isolated from diseased cerebellum tissue removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident. Patient history included Huntington's disease, emphysema, and tobacco abuse.
BRACNOK02	PSPORT1	This amplified and normalized library was constructed using RNA isolated from posterior cingulate tissue removed from an 85-year-old Caucasian female who died from myocardial infarction and retroperitoneal hemorrhage. Pathology indicated atherosclerosis, moderate to severe, involving the circle of Willis, middle cerebral, basilar and vertebral arteries; infarction, remote, left dentate nucleus; and amyloid plaque deposition consistent with age. There was mild to moderate leptomeningeal fibrosis, especially over the convexity of the frontal lobe. There was mild generalized atrophy involving all lobes. The white matter was mildly thinned. Cortical thickness in the temporal lobes, both maximal and minimal, was slightly reduced. The substantia nigra pars compacta appeared mildly depigmented. Patient history included COPD, hypertension, and recurrent deep venous thrombosis. 6.4 million independent clones from this amplified library were normalized in one round using conditions
		adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791.
BRAINOR03	PBK-CMV	This random primed library was constructed using pooled cDNA from two donors. cDNA was generated using mRNA isolated from brain tissue removed from a Caucasian male fetus (donor A) who was stillborn with a hypoplastic left heart at 23 weeks' gestation and from brain tissue removed from a Caucasian male fetus (donor B), who died at 23 weeks' gestation from premature birth. Serologies were negative for both donors and family history for donor B included diabetes in the mother.
MYEPUNF01	pRARE	This 5' cap isolated full-length library was constructed using RNA isolated from an untreated K-562 cell line, derived from chronic myelogenous leukemia precursor cells removed from a 53-year-old female.
MYEPUNN01	pRARE	This normalized untreated K-562 cell line tissue library was constructed from independent clones from a K-562 cell line library. Starting RNA was made from an untreated K-562 cell line, derived from chronic myelogenous leukemia precursor cells removed from a 53-year-old female. The library was normalized in one round using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.

Table 6

Library	Vector	Library Description
NERDTDN03	pINCY	This normalized dorsal root ganglion tissue library was constructed from 1.05 million independent clones from a dorsal root ganglion tissue library. Starting RNA was made from dorsal root ganglion tissue removed from the cervical spine of a 32-year-old Caucasian male who died from acute pulmonary edema, acute bronchopneumonia, bilateral pleural effusions, pericardial effusion, and malignant lymphoma (natural killer cell type). The patient presented with pyrexia of unknown origin, malaise, fatigue, and gastrointestinal bleeding. Patient history included probable cytomegalovirus infection, liver congestion, and steatosis, splenomegaly, hemorrhagic cystitis, thyroid hemorrhage, respiratory failure, pneumonia of the left lung, natural killer cell lymphoma of the pharynx, Bell's palsy, and tobacco and alcohol abuse. Previous surgeries included colonoscopy, closed colon biopsy, adenotonsillectomy, and nasopharyngeal endoscopy and biopsy. Patient medications included Diflucan (fluconazole), Deltasone (prednisone),
		hydrocodone, Lortab, Alprazolam, Reazodone, ProMace-Cytabom, Etoposide, Cisplatin, Cytarabine, and dexamethasone. The patient received radiation therapy and multiple blood transfusions. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
OVARDIN02	pINCY	This normalized ovarian tissue library was constructed from 5.76 million independent clones from an ovary library. Starting RNA was made from diseased ovarian tissue removed from a 39-year-old Caucasian female during total abdominal hysterectomy, bilateral salpingo-oophorectomy, dilation andcurettage, partial colectomy, incidental appendectomy, and temporary colostomy. Pathology indicated the right and left adnexa, mesentery and muscularis propria of the sigmoid colon were extensively involved by endometriosis. Endometriosis also involved the anterior and posterior serosal surfaces of the uterus and the cul-de-sac. The endometrium was proliferative. Pathology for the associated tumor tissue indicated multiple (3 intramural, 1 subserosal) leiomyomata. The patient presented with abdominal pain and infertility. Patient history included scoliosis. Family history included hyperlipidemia, benign hypertension, atherosclerotic coronary artery disease, depressive disorder, brain cancer, and type II diabetes. The
		library was normalized in two rounds using conditions adapted from Soares et al., PNAS(1994) 91:9228 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48-hours/round) reannealing hybridization was used.

Table 6

Library	Vector	Library Description
PITUNON01	pINCY	This normalized pituitary gland tissue library was constructed from 6.92 million independent clones from a pituitary gland tissue library. Starting RNA was made from pituitary gland tissue removed from a 55-year-old male who died from chronic obstructive pulmonary disease. Neuropathology indicated there were no gross abnormalities, other than mild ventricular enlargement. There was no apparent microscopic abnormality in any of the neocortical areas examined, except for a number of silver positive neurons with apical dendrite staining, particularly in the frontal lobe. The significance of this was undetermined. The only other microscopic abnormality was that there was prominent silver staining with some swollen axons in the CA3 region of the anterior and posterior hippocampus. Microscopic sections of the cerebellum revealed mild Bergmann's gliosis in the Purkinje cell layer. Patient history included schizophrenia. The library was normalized in two rounds using conditions (1994) 91:9228-9232 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
SININOT04	pINCY	Library was constructed using RNA isolated from diseased ileum tissue obtained from a 26-year-old Caucasian male during a partial colectomy, permanent colostomy, and an incidental appendectomy. Pathology indicated moderately to severely active Crohn's disease. Family history included enteritis of the small intestine.
THYMNOE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from thymus tissue removed from a 2-year-old Caucasian female during a thymectomy and patch closure of left atrioventricular fistula. Pathology indicated there was no gross abnormality of the thymus. The patient presented with congenital heart abnormalities. Patient history included double inlet left ventricle and a rudimentary right ventricle, pulmonary hypertension, cyanosis, subaortic stenosis, seizures, and a fracture of the skull base. Patient medications included Lasix and Captopril. Family history included reflux neuropathy in the mother.
UTREDMF02	PCMV-ICIS	This full-length enriched library was constructed using 1.5 micrograms of polyA RNA isolated from endometrial tissue removed from a 32-year-old female. The endometrium was in secretory phase.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value = 1.0E-8 or less; Full Length sequences: Probability value = 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value = 1.06E-6; Assembled ESTs: fasta Identity = 95% or greater and Match length = 200 bases or greater; fastx E value = 1.0E-8 or less; Full Length sequences: fastx score = 100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value = 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM, INCY, SMART or TIGRFAM hits: Probability value = 1.0E-3 or less; Signal peptide hits: Score = 0 or greater
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score ≥ GCG-specified "HIGH" value for that particular Prosite motif. Generally, score = 1.4-2.1.

Table 7

Program	Description	Reference	Parameter Threshold
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score = 120 or greater; Match length = 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score = 3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
16	7526185	125901F1	SNP00047602	255	3028	T	C	T	noncoding	0.61	0.47	0.61	0.61
16	7526185	1553407H1	SNP00155225	129	1488	T	T	C	noncoding	n/a	n/a	n/a	n/a
16	7526185	2197671T6	SNP00155225	216	1533	T	T	C	noncoding	n/a	n/a	n/a	n/a
16	7526185	6723530H1	SNP00051188	262	2713	A	G	A	noncoding	n/a	n/a	n/a	n/a
16	7526185	829638T6	SNP00155225	181	1582	C	T	C	noncoding	n/a	n/a	n/a	n/a
17	7526192	1208904H1	SNP00062572	151	2429	G	G	A	noncoding	n/a	n/a	n/a	n/d
17	7526192	1223444H1	SNP00098139	99	2262	C	C	T	noncoding	n/d	n/a	n/a	n/a
17	7526192	1231274R6	SNP00115694	8	1018	C	C	T	S113	n/a	n/a	n/a	n/a
17	7526192	1341206H1	SNP00068492	33	1909	A	A	G	noncoding	n/d	n/a	n/a	n/a
17	7526192	1405367T6	SNP00062572	57	2446	G	G	A	noncoding	n/a	n/a	n/a	n/a
17	7526192	1405367T6	SNP00098139	224	2279	C	C	T	noncoding	n/d	n/a	n/a	n/d
17	7526192	1417137T6	SNP00062572	34	2482	G	G	A	noncoding	n/a	n/a	n/a	n/a
17	7526192	1553058T6	SNP00062572	51	2453	G	G	A	noncoding	n/a	n/a	n/a	n/a
17	7526192	1678219T6	SNP00098139	232	2271	C	C	T	noncoding	n/d	n/a	n/a	n/d
17	7526192	1722718F6	SNP00068491	24	1686	C	C	T	noncoding	n/a	n/a	n/a	n/a
17	7526192	1722718F6	SNP00068492	249	1911	A	A	G	noncoding	n/d	n/a	n/a	n/a
17	7526192	1722718H1	SNP00068491	24	1683	C	C	T	noncoding	n/a	n/a	n/a	n/a
17	7526192	2997552T6	SNP00062572	180	2342	G	G	A	noncoding	n/a	n/a	n/a	n/a
17	7526192	2997552T6	SNP00098139	347	2177	C	C	T	noncoding	n/d	n/a	n/a	n/d
17	7526192	7674218H2	SNP00068492	342	1908	A	A	G	noncoding	n/d	n/a	n/a	n/a
18	7526193	1328791H1	SNP00057788	217	716	G	G	T	noncoding	n/a	n/a	n/a	n/a
18	7526193	4291033F6	SNP00142508	164	1539	A	A	G	N232	n/a	n/a	n/a	n/a
18	7526193	4291033F6	SNP00142509	191	1566	A	A	G	T241	n/a	n/a	n/a	n/a
18	7526193	7217965H1	SNP00118120	200	145	G	A	G	noncoding	n/d	n/a	n/a	n/a
18	7526193	7760201H1	SNP00057788	118	715	G	G	T	noncoding	n/a	n/a	n/a	n/a
19	7526196	1216956H1	SNP00006796	60	4315	A	A	G	noncoding	0.98	n/a	n/a	n/a
19	7526196	1224406H1	SNP00124328	162	2487	A	A	G	noncoding	n/d	n/a	n/a	n/a
19	7526196	1436210H1	SNP00006288	103	3533	G	G	A	noncoding	0.97	n/a	n/a	n/a
19	7526196	1438205H1	SNP00124330	102	3121	A	A	G	noncoding	n/d	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
19	7526196	1555235H1	SNP00006287	133	3401	C	T	C	noncoding	0.34	n/a	n/a	n/a
19	7526196	1597263F6	SNP00124329	113	2824	A	A	G	noncoding	n/d	n/d	n/d	n/d
19	7526196	1669032H1	SNP00124327	46	65	C	T	C	noncoding	n/a	n/a	n/a	n/a
19	7526196	2555446F6	SNP00068980	190	1856	A	A	G	noncoding	n/d	n/d	n/d	n/d
19	7526196	2555446H1	SNP00068980	190	1854	A	A	G	noncoding	n/d	n/d	n/d	n/d
19	7526196	3337906H1	SNP00153438	134	2187	G	G	A	noncoding	n/a	n/a	n/a	n/a
19	7526196	3643184H1	SNP00068979	155	1594	C	C	T	noncoding	n/a	n/a	n/a	n/a
19	7526196	6754284H1	SNP00068978	420	1195	A	A	G	noncoding	n/a	n/a	n/a	n/a
19	7526196	7703764J1	SNP00124330	162	3128	A	A	G	noncoding	n/d	n/a	n/a	n/a
19	7526196	7753868H1	SNP00124329	510	2852	A	A	G	noncoding	n/d	n/d	n/d	n/d
19	7526196	7753868H1	SNP00124330	213	3149	A	A	G	noncoding	n/d	n/a	n/a	n/a
19	7526196	8598525H1	SNP00006287	238	3428	T	T	C	noncoding	0.34	n/a	n/a	n/a
19	7526196	8598525H1	SNP00006288	370	3560	G	G	A	noncoding	0.97	n/a	n/a	n/a
20	7526198	1216956H1	SNP00006796	60	4122	A	A	G	K1309	0.98	n/a	n/a	n/a
20	7526198	1224406H1	SNP00124328	162	2284	A	A	G	K696	n/d	n/a	n/a	n/a
20	7526198	1436210H1	SNP00006288	103	3331	G	G	A	A1045	0.97	n/a	n/a	n/a
20	7526198	1438205H1	SNP00124330	102	2918	A	A	G	T908	n/d	n/a	n/a	n/a
20	7526198	1555235H1	SNP00006287	133	3199	C	T	C	S1001	0.34	n/a	n/a	n/a
20	7526198	1597263F6	SNP00124329	113	2621	A	A	G	N809	n/d	n/d	n/d	n/d
20	7526198	1669032H1	SNP00124327	46	65	C	T	C	noncoding	n/a	n/a	n/a	n/a
20	7526198	1806969T6	SNP00029581	329	4272	C	C	T	noncoding	n/a	n/a	n/a	n/a
20	7526198	1922794H1	SNP00092542	84	4217	G	G	A	D1341	n/a	n/a	n/a	n/a
20	7526198	2005750H1	SNP00029581	61	4270	C	C	T	noncoding	n/a	n/a	n/a	n/a
20	7526198	2189973H1	SNP00136926	25	4147	C	C	T	A1317	n/a	n/a	n/a	n/a
20	7526198	2555446F6	SNP00068980	190	1655	A	A	G	N487	n/d	n/d	n/d	n/d
20	7526198	2555446H1	SNP00068980	190	1653	A	A	G	Y486	n/d	n/d	n/d	n/d
20	7526198	2936740H1	SNP00006289	104	3928	C	C	T	T1244	0.91	n/a	n/a	n/a
20	7526198	3337906H1	SNP00153438	134	1984	G	G	A	R596	n/a	n/a	n/a	n/a
20	7526198	3643184H1	SNP00068979	155	1393	C	C	T	S399	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
20	7526198	6754284H1	SNP00068978	420	994	A	A	G	G266	n/a	n/a	n/a	n/a
20	7526198	7703764J1	SNP00124330	162	2925	A	A	G	Q910	n/d	n/a	n/a	n/a
20	7526198	7753868H1	SNP00124329	510	2649	A	A	G	D818	n/d	n/d	n/d	n/d
20	7526198	7753868H1	SNP00124330	213	2946	A	A	G	H917	n/d	n/a	n/a	n/a
20	7526198	8598525H1	SNP00006287	238	3226	T	T	C	D1010	0.34	n/a	n/a	n/a
20	7526198	8598525H1	SNP00006288	370	3358	G	G	A	T1054	0.97	n/a	n/a	n/a
21	7526208	1236920F1	SNP00033469	330	3805	C	C	T	noncoding	n/a	n/a	n/a	n/a
21	7526208	1284901H1	SNP00013862	33	2502	C	C	G	Q454	0.91	n/a	n/a	n/a
21	7526208	1915448H1	SNP00003491	119	5753	A	A	G	noncoding	n/a	n/a	n/a	n/a
21	7526208	2528372H1	SNP00053975	148	2683	C	C	T	noncoding	n/a	n/a	n/a	n/a
21	7526208	2681418H1	SNP00053974	116	2185	T	T	C	L348	n/d	n/a	n/a	n/a
21	7526208	2749684F6	SNP00153340	187	1794	T	T	C	W218	n/a	n/a	n/a	n/a
21	7526208	3331418H1	SNP00132658	181	2866	T	T	C	noncoding	n/a	n/a	n/a	n/a
21	7526208	3693823H1	SNP00053972	67	1115	C	C	T	noncoding	n/d	n/a	n/a	n/a
21	7526208	3967421F6	SNP00033469	163	3806	C	C	T	noncoding	n/a	n/a	n/a	n/a
21	7526208	5055874H1	SNP00113323	139	5405	T	T	G	noncoding	n/a	n/a	n/a	n/a
21	7526208	6449431H1	SNP00053973	480	1969	T	C	T	V276	n/a	n/a	n/a	n/a
22	7526212	2749684F6	SNP00153340	187	1794	T	T	C	W218	n/a	n/a	n/a	n/a
22	7526212	3693823H1	SNP00053972	67	1115	C	C	T	noncoding	n/d	n/a	n/a	n/a
22	7526212	6449431H1	SNP00053973	480	1969	T	C	T	V276	n/a	n/a	n/a	n/a
23	7526213	1430148F6	SNP00014900	123	3197	C	G	C	noncoding	0.04	n/a	n/a	n/a
23	7526213	2113230H1	SNP00044591	151	6811	C	C	T	noncoding	n/a	n/a	n/a	n/a
23	7526213	2113230R6	SNP00044591	151	6814	C	C	T	noncoding	n/a	n/a	n/a	n/a
23	7526213	2556574H1	SNP00139931	137	2468	A	A	G	noncoding	n/a	n/a	n/a	n/a
23	7526213	2987033F6	SNP00153180	352	6398	C	C	T	noncoding	n/a	n/a	n/a	n/a
23	7526213	3844660H1	SNP00014900	64	3198	C	G	C	noncoding	0.04	n/a	n/a	n/a
23	7526213	712904R6	SNP00139930	72	2210	A	G	A	noncoding	n/a	n/a	n/a	n/a
24	7526214	1430148F6	SNP00014900	123	3134	C	G	C	noncoding	0.04	n/a	n/a	n/a
24	7526214	2113230H1	SNP00044591	151	6746	C	C	T	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
24	7526214	2113230R6	SNP00044591	151	6749	C	C	T	noncoding	n/a	n/a	n/a	n/a
24	7526214	2556574H1	SNP00139931	137	2403	A	A	G	noncoding	n/a	n/a	n/a	n/a
24	7526214	2987033F6	SNP00153180	352	6333	C	C	T	noncoding	n/a	n/a	n/a	n/a
24	7526214	3844660H1	SNP00014900	64	3135	C	G	C	noncoding	0.04	n/a	n/a	n/a
25	7526228	1485690T6	SNP00066816	319	2716	G	G	A	noncoding	n/d	n/d	0.96	n/d
25	7526228	1835249H1	SNP00066816	8	2712	G	G	A	noncoding	n/d	n/d	0.96	n/d
25	7526228	2169542T6	SNP00066816	241	2740	G	G	A	noncoding	n/d	n/d	0.96	n/d
25	7526228	2536771H1	SNP00136441	18	234	G	G	A	noncoding	n/a	n/a	n/a	n/a
25	7526228	2805663T6	SNP00066816	251	2787	G	G	A	noncoding	n/d	n/d	0.96	n/d
25	7526228	510019T6	SNP00066816	194	2713	G	G	A	noncoding	n/d	n/d	0.96	n/d
26	7526246	1294154H1	SNP00068998	123	2362	A	A	G	noncoding	n/a	n/a	n/a	n/a
26	7526246	1545488H1	SNP00068997	78	2115	C	C	G	noncoding	n/a	n/a	n/a	n/a
26	7526246	280325T6	SNP00068997	58	2182	C	C	G	noncoding	n/a	n/a	n/a	n/a
26	7526246	4407121H1	SNP00041996	224	2946	A	A	G	noncoding	n/a	n/a	n/a	n/a
26	7526246	7621966J1	SNP00068997	165	2116	C	C	G	noncoding	n/a	n/a	n/a	n/a
26	7526246	7751044H1	SNP00068998	486	2365	A	A	G	noncoding	n/a	n/a	n/a	n/a
27	7526258	1348638F6	SNP00076027	241	328	G	G	C	G74	n/d	n/d	n/d	n/d
27	7526258	1348638F6	SNP00132757	63	150	A	A	G	R14	n/a	n/a	n/a	n/a
27	7526258	1444773H1	SNP00037439	67	402	G	G	C	L98	n/a	n/a	n/a	n/a
27	7526258	1897166H1	SNP00043983	191	1194	T	T	C	P362	n/a	n/a	n/a	n/a
27	7526258	2770947H1	SNP00154171	21	1113	C	T	C	D335	n/a	n/a	n/a	n/a
27	7526258	3143852H1	SNP00037440	73	1151	T	C	T	L348	n/d	n/d	n/d	n/d
27	7526258	3143852H1	SNP00111294	30	1108	C	C	T	L334	1	n/d	n/d	n/d
28	7526311	1649261F6	SNP00019740	300	2179	T	T	C	noncoding	n/a	n/a	n/a	n/a
28	7526311	1649261T6	SNP00019740	252	2259	T	T	C	noncoding	n/a	n/a	n/a	n/a
28	7526311	268900T6	SNP00019740	257	2257	T	T	C	noncoding	n/a	n/a	n/a	n/a
28	7526311	2745158H1	SNP00058093	32	1282	T	T	C	noncoding	0.82	0.9	0.98	0.92
28	7526311	2745158H1	SNP00114001	97	1347	G	T	G	noncoding	n/a	n/a	n/a	n/a
28	7526311	2921293T6	SNP00019740	268	2235	T	T	C	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
28	7526311	8011285H1	SNP00125603	438	540	C	C	T	A129	n/a	n/a	n/a	n/a
29	7526315	058064H1	SNP00003740	186	1775	A	A	G	noncoding	n/a	n/a	n/a	n/a
29	7526315	1004004H1	SNP00012539	79	5335	T	C	T	noncoding	n/a	n/a	n/a	n/a
29	7526315	1004004H1	SNP00012540	191	5447	G	A	G	noncoding	0.71	0.63	0.86	0.64
29	7526315	1004004H1	SNP00045700	205	5461	C	C	T	noncoding	n/a	n/a	n/a	n/a
29	7526315	1330039H1	SNP00045701	48	5601	G	G	C	noncoding	n/a	n/a	n/a	n/a
29	7526315	1363254H1	SNP00022215	65	2848	A	A	C	noncoding	n/d	n/a	n/a	n/a
29	7526315	1377277F1	SNP00012538	87	4783	C	C	T	noncoding	n/a	n/a	n/a	n/a
29	7526315	1675313F6	SNP00028237	82	3316	T	T	C	noncoding	n/a	n/a	n/a	n/a
29	7526315	1675313F6	SNP00028238	133	3367	A	G	A	noncoding	n/d	n/a	n/a	n/a
29	7526315	1675313T6	SNP00012538	16	4780	C	C	T	noncoding	n/a	n/a	n/a	n/a
29	7526315	1682961T7	SNP00045701	458	5606	G	G	C	noncoding	n/a	n/a	n/a	n/a
29	7526315	3003741H1	SNP00023889	140	215	T	T	G	G5	n/a	n/a	n/a	n/a
29	7526315	403838T6	SNP00028238	336	3387	G	G	A	noncoding	n/d	n/a	n/a	n/a
29	7526315	7622837J1	SNP00028237	242	3317	T	T	C	noncoding	n/a	n/a	n/a	n/a
29	7526315	7622837J1	SNP00028238	191	3368	G	G	A	noncoding	n/d	n/a	n/a	n/a
29	7526315	7625836H1	SNP00028237	72	3299	T	T	C	noncoding	n/a	n/a	n/a	n/a
29	7526315	7625836H1	SNP00028238	123	3348	G	G	A	noncoding	n/d	n/a	n/a	n/a
29	7526315	7752327H1	SNP00045701	517	5600	G	G	C	noncoding	n/a	n/a	n/a	n/a
30	7526442	1265917F1	SNP00149600	355	1648	T	T	C	noncoding	n/a	n/a	n/a	n/a
30	7526442	1382145F6	SNP00149600	469	1646	T	T	C	noncoding	n/a	n/a	n/a	n/a
30	7526442	1824201F6	SNP00066979	79	724	C	C	T	S139	n/d	n/d	n/a	n/d
30	7526442	2046231H1	SNP00114113	226	1004	G	G	C	noncoding	n/d	n/a	n/a	n/a
30	7526442	2744627F6	SNP00022802	98	167	C	C	G	noncoding	n/a	n/a	n/a	n/a
30	7526442	691185T6	SNP00149600	55	1722	T	T	C	noncoding	n/a	n/a	n/a	n/a
30	7526442	7622751H1	SNP00031991	177	769	C	C	T	noncoding	n/a	n/a	n/a	n/a